Development of new genome-informed genotyping tools for *Aphanomyces astaci*

Submitted by Diana Minardi to the University of Exeter

as a thesis for the degree of Doctor of Philosophy

in Biological Sciences

In May 2017

This thesis is available for Library use on the understanding that it is copyright material and that no quotation from the thesis may be published without proper acknowledgement.

I certify that all material in this thesis which is not my own work has been identified and that no material has previously been submitted and approved for the award of a degree by this or any other University.

Signature: ……..

Acknowledgements

I would like to sincerely thank my supervisors Dr Mark van der Giezen, Dr David Studholme, and Dr Birgit Oidtmann for their invaluable guidance and support during these 3 years (plus a bit) of PhD, both practically and in the writing of the thesis.

I would like to thank Cefas and the University of Exeter for funding the project and for giving me the opportunity to embrace this challenge and adventure.

I would like to thank all of the University of Exeter “Biocat” lab and office, with whom I’ve spent most of my coffee time. But especially Mirella (for the laughs and cries) and Simone (for the sweaty runs around campus). I would also like to thank Chiara for her constant support, encouragement, and beer.

Finally, I would like to thank Tony, for his patience, and my family, for their “overseas” support: I’ve finally finished “the shrimp” book!

Abstract

Aphanomyces spp. are water moulds, eukaryotic fungus-like organisms, belonging to the class Oomycota. This genus contains primary pathogens of plants and animals as well as opportunistic and saprotrophic species. One of the animal parasites (*A. astaci*) is the causal agent of the crayfish plague, a disease listed by the World Organisation for Animal Health (OIE). It is believed that *A. astaci* was first introduced into Italy from the US in the late 19th century and rapidly spread in Europe causing the decline of native crayfish. It currently threatens to wipe out the UK native white-clawed crayfish (*Austropotamobius pallipes*). Random amplified polymorphic DNA (RAPD-PCR) on pure isolates of *A. astaci* distinguished five genotypes (A, B, C, D, and E). This distinction proved to be a useful tool for epidemiological studies aimed at understanding the history and spread of the disease in Europe; furthermore, there are differences in virulence among genotypes. No discriminatory morphological or physiological characters are available and widely used markers such as the internal transcribed spacer (ITS), the divergent domains regions (D1-D2) of nuclear large subunit (LSU) rDNA, and cytochrome c oxidase subunit I (COI) also fail to discriminate between *A. astaci* genotypes. There are some practical drawbacks to genotype by the currently available genotyping methods. Whole-genome sequencing (WGS) was used to catalogue DNA single nucleotide variants and genotype-unique genomic regions that could be exploited as phylogenetic markers. These newly developed molecular markers were tested both on pure cultures and historical samples derived from outbreaks and carrier crayfish available in our laboratories, validating these genotyping methods, which represent new diagnostic tools aiding the detection and prevention of crayfish plague.

List of Contents

List of Tables

List of Figures

List of Abbreviations

°C	Celsius
µg	microgram
µl	microlitre
µm	micrometre
µM	micromolar
AFLP	amplified fragment length polymorphism
ag	attogram
BLAST	Basic Local Alignment Search Tool
bp	base pair
BUSCO	Benchmarking Universal Single-Copy Orthologs
BWA	Burrows-Wheeler Aligner
CAZys	carbohydrate-active enzymes
Cefas	Centre for Environment, Fisheries and Aquaculture Science
CHI1 to 3	chitinases groups 1 to 3
cm	centimetre
CRN	crinkler (from crinkling and necrosis)
COI	cytochrome oxidase subunit I
DC	cuticle from outbreak dead crayfish
DNA	deoxyribonucleic acid
EDTA	ethylenediaminetetraacetic acid
fg	femtogram
g	gram
gDNA	high molecular weight genomic DNA

GH20	glycoside hydrolases belonging to the family 20
GP	glucose-peptone media
IGV	Integrative Genomics Viewer
ITS	internal transcribed spacer (partial 18 S rDNA gene, ITS1 region, 5.8 S rDNA gene, ITS2 region and partial 28 S rDNA gene)
kbp	kilobase pair
LB	Luria-Bertani media
LOH	loss of heterozygosity
LSU	large subunit ribosomal DNA (partial 18 S rDNA gene, ITS1 region, 5.8 S rDNA gene, ITS2 region and partial 28 S rDNA gene)
LxLFLAK	leucine-x-leucine-phenylalanine-leucine-alanine-lysine, x as any amino acid, protein families
M	molar
MC	cuticle from outbreak moribund crayfish
mg	milligram
min	minute
ml	millilitre
mm	millimetre
mM	millimolar
mtDNA	mitochondrial DNA
NCBI	National Center for Biotechnology Information
ng	nanogram
NGS	next generation sequencing
OCC	Oomycetes Culture Collection

OIE	World Organisation for Animal Health
p	polymorphism level (number of variants/fragment length*100)
PCR	polymerase chain reaction
pg	picogram
qPCR	quantitative PCR
RAPD-PCR	random amplification of polymorphic DNA
RcA	semi-nested PCR reaction control (genotype A)
RcB	semi-nested PCR reaction control (genotype B)
RNA	ribonucleic acid
RxLR	arginine-x-leucine-arginine, x as any amino acid, protein families
RZE	repeated zoospore emergence
SDS	sodium dodecyl sulfate
sec	second
SNV	single nucleotide variant
SOB	super optimal broth
SpHtp1	*Saprolegnia parasitica* host-targeting protein 1
TAE	Tris base, acetic acid and EDTA electrophoresis buffer
USCOs	Universal Single-Copy Orthologs
UV	ultraviolet light
WGS	whole genome sequencing
WSSV	white-spot syndrome virus

1. Introduction

1.1 *Aphanomyces*, an overview

Oomycetes (class: Oomycota) are filamentous eukaryotic fungus-like organisms commonly known as water moulds due to their preference for aquatic environments. This class comprises saprotrophic species as well as plant, marine algae and animal pathogenic species (Figure 1.1), responsible of devastating economic losses in agriculture (*e.g. Phytophthora infestans*, causative agent of the late blight disease in potatoes and tomatoes), aquaculture (*e.g. Saprolegnia parasitica*, causing saprolegniasis in salmonids) and in wild populations (*e.g. Aphanomyces astaci*, causing the crayfish plague and involved in the decline of European indigenous crayfish species; *Phytophthora ramorum*, causing the sudden oak death disease in oak trees) (Alderman *et al*., 1984; Duncan, 1999; Fry and Goodwin, 1997; Grenville-Briggs *et al*., 2011; Rizzo *et al*., 2002; van den Berg *et al*., 2013).

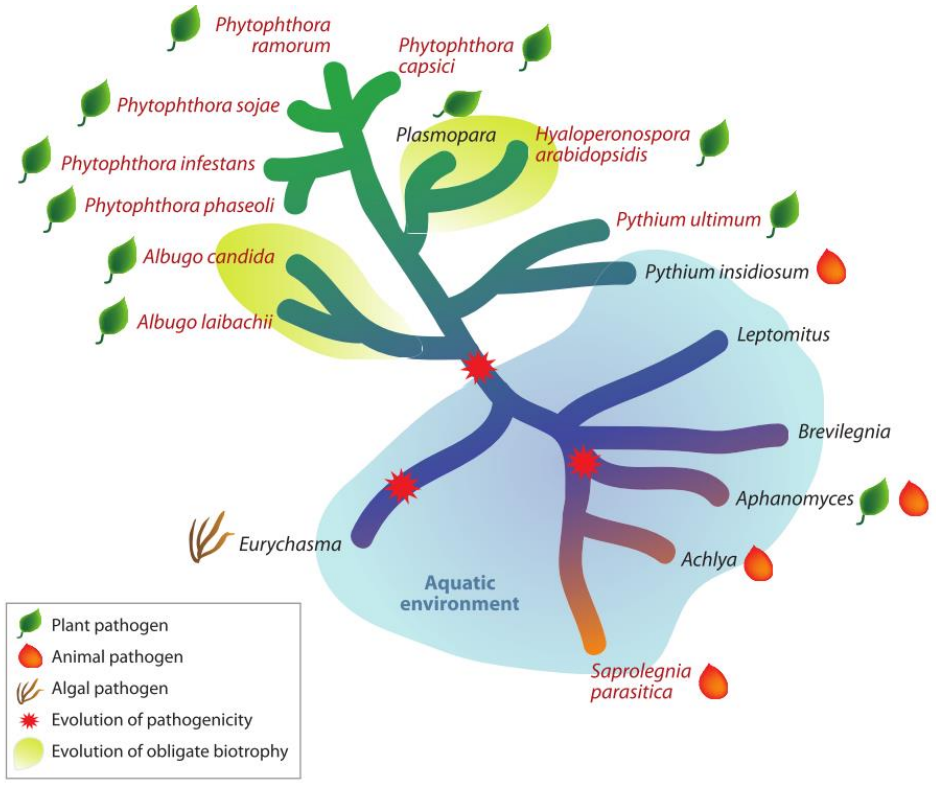


Figure 1.1 Schematic representation of plant and animal pathogenic oomycetes inferred from the evolution of pathogenicity and obligate biotrophy. Species in red: whole genome sequences available. Image from Jiang *et al*. (2013).

Differently to other oomycetes, the genus *Aphanomyces* includes species that successfully colonised marine, freshwater, and terrestrial environments, comprising plant and animal primary pathogens (Figure 1.1, green and red marks next to *Aphanomyces* genus) but also opportunistic and saprotrophic species that can grow on decaying organic materials (Diéguez-Uribeondo *et al.*, 2006). Phylogenetic analysis based on internal transcribed spacer (partial 18 S rDNA gene, ITS1 region, 5.8 S rDNA gene, ITS2 region and partial 28 S rDNA gene, ITS) (Figure 1.2), show the distinction of these three main lineages of *Aphanomyces*: a plant parasitic lineage, a saprotrophic/opportunistic lineage and a predominantly animal parasitic one (Diéguez-Uribeondo *et al.*, 2009). This last clade includes animal parasites such as *A. astaci* and *A. invadans*, species isolated from animals, such as *A. frigidophilus*, and saprotrophic species, such as *A. stellatus*. Among the animal parasites, *A. astaci* is responsible for crayfish plague, a disease listed by the World Organisation for Animal Health (OIE) that caused the decline of European crayfish and extensive damage to both wild and cultured European crayfish stocks (Alderman, 1996; OIE, 2016a; Söderhäll and Cerenius, 1999).

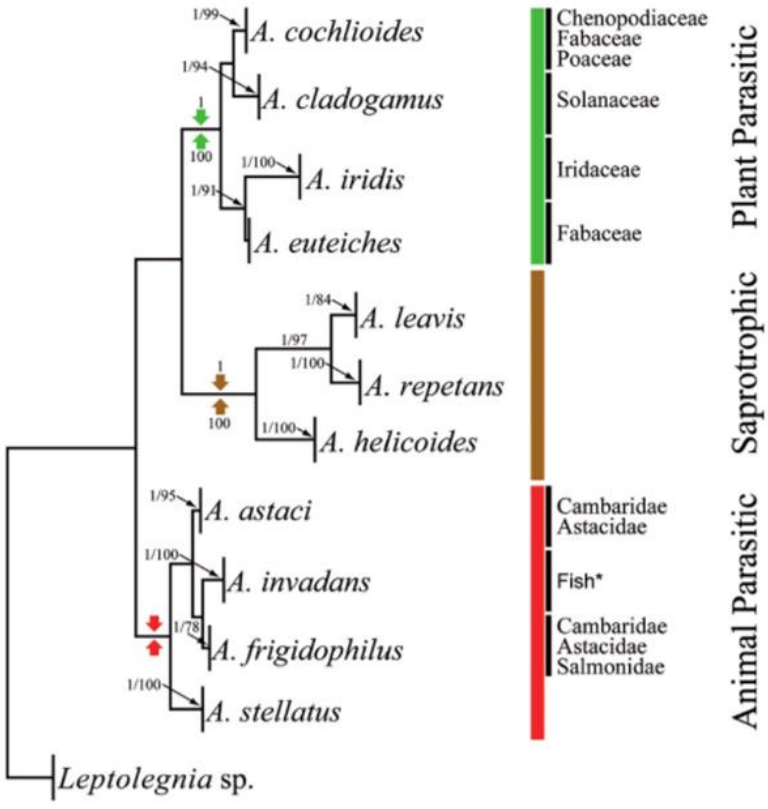


Figure 1.2 ITS based phylogenetic analysis of the *Aphanomyces* genus, showing the different *Aphanomyces* lineages. The analysis includes plant and animal parasites and opportunistic/saprotrophic species. Green, brown, and red arrows indicate the three major clades. Image from Rezinciuc *et al.* (2015).

1.1.1 Life cycle of *Aphanomyces* species

As with other oomycetes, the life cycles of *Aphanomyces* include a sexual and an asexual stage which both take place in the hyphae of the vegetative mycelium (Diéguez-Uribeondo *et al.*, 2006). The main functions of the first type of reproduction are to assure genetic variability and to produce resistant resting stages that can ensure the survival of the parasite, while the functions for the asexual stage are the production and release of motile units to ensure rapid dispersal of the parasites in the environment (Diéguez-Uribeondo *et al.*, 2006). In the *Aphanomyces* genus, differences in the type of reproduction undertaken by different species have been observed. In general, plant parasitic species (*i.e. A. cladogamus, A. cochlioides, A. euteiches* and *A. iridis*) and saprotrophic species (*i.e. A. helicoides, A. leavis and A. stellatus*) undertake both sexual and asexual reproduction. On the opposite, in animal parasitic species the sexual stage is rare (*i.e. A. frigidophilus*) or has never been observed under laboratory conditions (*i.e. A. astaci* and *A. invadans*). In animal parasitic species the reproduction is mainly asexual and characterised by the formation of biflagellate motile zoospores which represent the infective units (Diéguez-Uribeondo *et al.*, 2009; Kitancharoen and Hatai, 1997). Sexual reproduction starts with the formation of antheridium (male) and oogonium (female) gametangia, which are specialised organs in which gametes are produced (Figure 1.3, sexual reproduction). These sexual structures fuse together by a fertilisation tube and produce oospores, a resistant stage that can live several years, or germinate, resulting in a new mycelium (Phillips *et al.*, 2008). The asexual stage starts when there is a significant reduction of nutrients in the environment surrounding the water mould, which triggers the formation of the sporangia (Figure 1.3, asexual reproduction). These structures are made of hyphal wall and contain cytoplasm, which split into individual motile biflagellate spores, called primary zoospores (Tiffney, 1939). In the sporangia, the primary zoospores are distributed in a single row and actively move towards the tip of the hyphae, where they are extruded. As soon as they reach the external environment, the primary zoospores undergo a process called encystment in which they round up, form a cell wall and adhere to each other, forming a cluster (or ball) of primary cysts, characteristic of the genus *Aphanomyces* (Buller, 2014). After resting, the primary cyst releases a biflagellate free swimming secondary

zoospore able to locate a suitable host to infect, by attaching and progressively encysting, forming a secondary cyst, germinating and originating new hyphae and new mycelium (Diéguez-Uribeondo *et al.*, 2006). The period of time that the secondary zoospores are motile varies, depending on environmental conditions, such as temperature (Alderman, 2003) and the presence of a host or a suitable substrate (Lilley *et al.*, 1998). In case secondary zoospores encyst on an unsuitable substrate, the secondary cysts can release a new generation of free-swimming zoospores, increasing the chances to subsequently locate a host or a suitable substrate to germinate. This feature is called repeated zoospore emergence (RZE), or polyplanetism. This feature is a peculiarity of parasitic *Aphanomyces* (*i.e. A. astaci, A. cochlioides* and *A. euteiches*), while other saprophytes species tend to germinate immediately after encystment (*i.e. A. helicoides, A. laevis* and *A. stellatus*) (Cerenius and Söderhäll, 1985; Royo *et al.*, 2004).

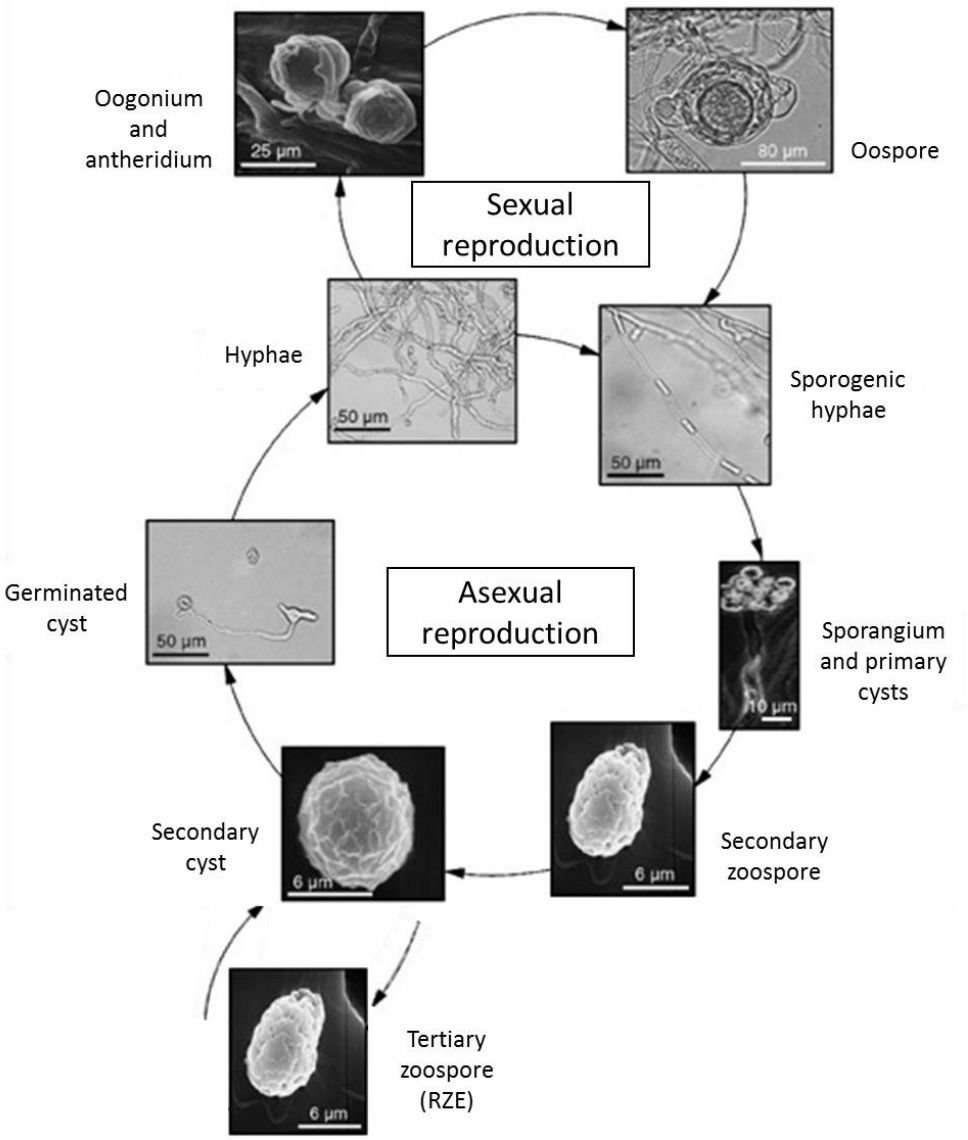


Figure 1.3 Schematic representation of sexual and asexual reproductions of *Aphanomyces* species. Sexual reproduction is characterised by the formation of antheridium (male) and oogonium (female) gametangia which fuse producing oospores. Asexual reproduction is marked by the formation of primary zoospores in the hyphae, which move towards the hyphal tip forming a sporangium. Once extruded, the primary zoospores encyst, forming the primary cysts. These cysts, release a biflagellate free swimming secondary zoospore able to locate a suitable host to infect, by attaching and progressively encysting, forming a secondary cyst, germinating, and originating new hyphae. If the secondary zoospores encyst on an unsuitable substrate, the secondary cysts can release a new generation of free-swimming zoospores, feature called repeated zoospore emergence (RZE). Image adapted from Gaulin *et al.* (2007).

1.1.2 Taxonomy of the genus *Aphanomyces* and isolation difficulties

The description of *Aphanomyces* species is based on the observation of sexual structures and, at present, this genus comprises 41 species with designated type species *A. stellatus* de Bary 1860 (Diéguez-Uribeondo *et al.*, 2009; Royo *et al.*, 2004; Takuma *et al.*, 2013). However, good descriptions of these structures in this genus are limited and the sexual phase is missing in animal pathogenic *Aphanomyces*, such as *A. astaci* and *A. invadans*, and therefore the exact number of validly described species is uncertain (Diéguez-Uribeondo *et al.*, 2009). Moreover, as there are no reliable morphological and physiological characteristics that allow the discrimination of animal parasitic *Aphanomyces* species from other saprotrophic/opportunistic aquatic *Aphanomyces*, the identification of these parasites is challenging (Royo *et al.*, 2004). In the past, this identification was often based only on host range, disease history and observation of non-septated oomycete hyphae in lesions (Edgerton *et al.*, 2004; Lilley *et al.*, 1998; Taugbøl *et al.*, 1993). At present, molecular techniques have helped the identification of *Aphanomyces* species by amplification and sequencing of ITS regions from pure cultures. However, these oomycetes are difficult to isolate into pure culture. When a successful isolation is obtained, cultures are hard to maintain alive and uncontaminated in laboratory conditions and, in many cases, reference strains are missing (Diéguez-Uribeondo *et al.*, 2009; Oidtmann *et al.*, 1999; Onions, 1971). Only recently, new molecular techniques to detect *A. astaci* have been developed to improve crayfish plague diagnosis from fresh or preserved infected crayfish samples, bypassing the isolation of the pathogen in pure culture, such as the microsatellites fingerprinting and the quantitative PCR (qPCR) (Grandjean *et al.*, 2014; Oidtmann *et al*., 2006; Vrålstad *et al.*, 2009).

1.2 *Aphanomyces astaci* and crayfish plague

1.2.1 Epidemiological summary of crayfish plague

Aphanomyces astaci is a specialised parasite of freshwater crayfish, heavily involved in the decline of European indigenous crayfish species (Alderman *et al.*, 1984; Unestam, 1972). The origin of the infection is still unknown but it is most likely that the parasite was first accidentally introduced in Italy at the end of the 19th century with the import of North American crayfish species, which can carry the parasites in the melanised cuticle as a chronic infection (Unestam, 1972). The firsts recorded European crayfish (*Astacus astacus* – noble crayfish) mass mortalities were from North Italy as described by Cornalia (1860), Martinati (1862), and Ninni (1865) [cited by Alderman *et al.* (1984)]. Unfortunately, the causes of these mortalities were not supported by direct examination of the pathogen involved. However, the extent of the mortalities and the severity of the disease reported suggests the presence of a highly infectious disease, such as crayfish plague (Alderman, 1996). From the first outbreak, other crayfish mass mortalities were recorded across Europe and, in only 140 years, the disease rapidly spread through central Europe, reaching the Balkan Peninsula, the Black Sea, Turkey, and Russia in the east, Spain in the south and entering Northern Europe, and the British Isles (Figure 1.4) (Alderman, 1996; Alderman *et al.*, 1984; Unestam, 1972). Thanks to the extensive records and studies on crayfish plague in Europe, it is possible to distinguish two epidemic waves: a first one, which ranges from the first introduction of the pathogen to the 1990's (Alderman, 1996), and a second one, which ranges from the 1980's onwards (Jussila *et al.*, 2016), including an overlapping period between the 1980's to the 1990's (Kozubíková *et al.*, 2008). This division of crayfish plague in waves is also supported by the genetic characterization of *A. astaci* pure culture isolates, which are now divided in five genotypes (or genogroups). For a comprehensive discussion on *A. astaci* genotypes, the reader is directed to section 1.2.5.

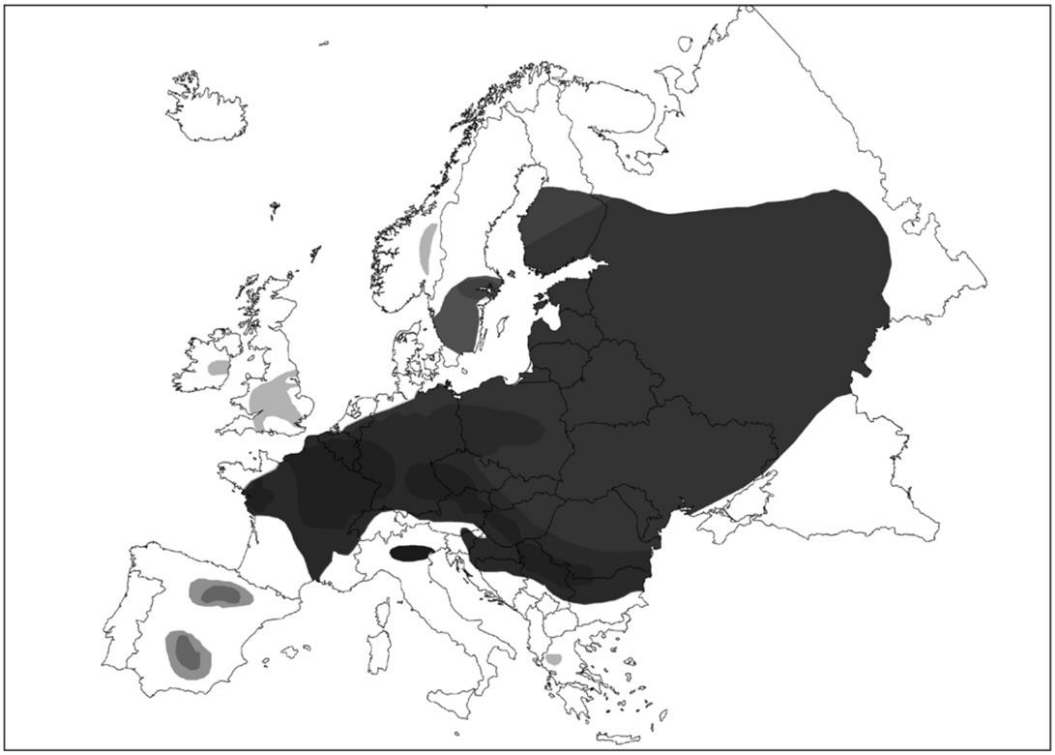


Figure 1.4 Geographical spread of crayfish plague in Europe from 1860 to 1995. Each grey shade represents a different time. From the darkest of the first introduction to the lightest: 1860–1869, 1870–1879, 1880–1889, 1890–1899, 1900–1909, 1910–1929, 1950–1969, 1970–1979, and 1980–1995. Figure from Jussila *et al*. (2016).

Recent European crayfish mortalities caused by *A. astaci* have been identified in north, central and south Italy (Caprioli *et al*., 2013; Marino *et al*., 2014; Pretto *et al.*, 2014), Czech Republic (Kozubíková *et al*., 2008), Croatia (Maguire *et al*., 2016), Slovenia (Kušar *et al*., 2013), Finland (Viljamaa-Dirks *et al.,* 2013) and, lately, in 2017 in Ireland (http://www.biodiversityireland.ie/crayfish-plague-biosecurity/).

1.2.2 Economic and environmental impact of crayfish plague

The economic impact of crayfish plague in the early stages of introduction in Europe is difficult to estimate as data are lacking. However, being a cheap source of protein especially for the lower class of society, crayfish have been used for human consumption for centuries and therefore the impact of crayfish mass mortalities and catchment declines could have had a significant effect on human communities (Cornalia, 1860; Patoka *et al*., 2016). In the present days, besides being considered a delicacy in Europe, freshwater crayfish have strong

cultural and recreational values, especially in Scandinavia where they are traditionally consumed throughout August in crayfish festivals (Edsman, 2004; Jussila *et al.,* 2015). At present, five species of crayfish belonging to two genera are considered native to Europe: the noble crayfish (*Astacus astacus*), which is found in Central and Northern Europe (Figure 1.5 A); the narrow-clawed crayfish (*Astacus leptodactylus*), which is found abundantly in Eastern Europe and the Middle East, but recently spread to other European countries (Figure 1.5 B); the thick-clawed crayfish (*Astacus pachypus*), restricted to the Black and Caspian Seas; the white-clawed crayfish (*Austropotamobius pallipes*) found in Southern Europe and the British Isles (Figure 1.5 C); the stone crayfish (*Austropotamobius torrentium*), found in the Alps and Balkans (Figure 1.5 D) (Kouba *et al*., 2014). Figure 1.5 (A-D) displays the geographical spread of the four major European crayfish species, with presumed native range (highlighted in red) and the most recent reported populations (black circles) (Kouba *et al*., 2014).

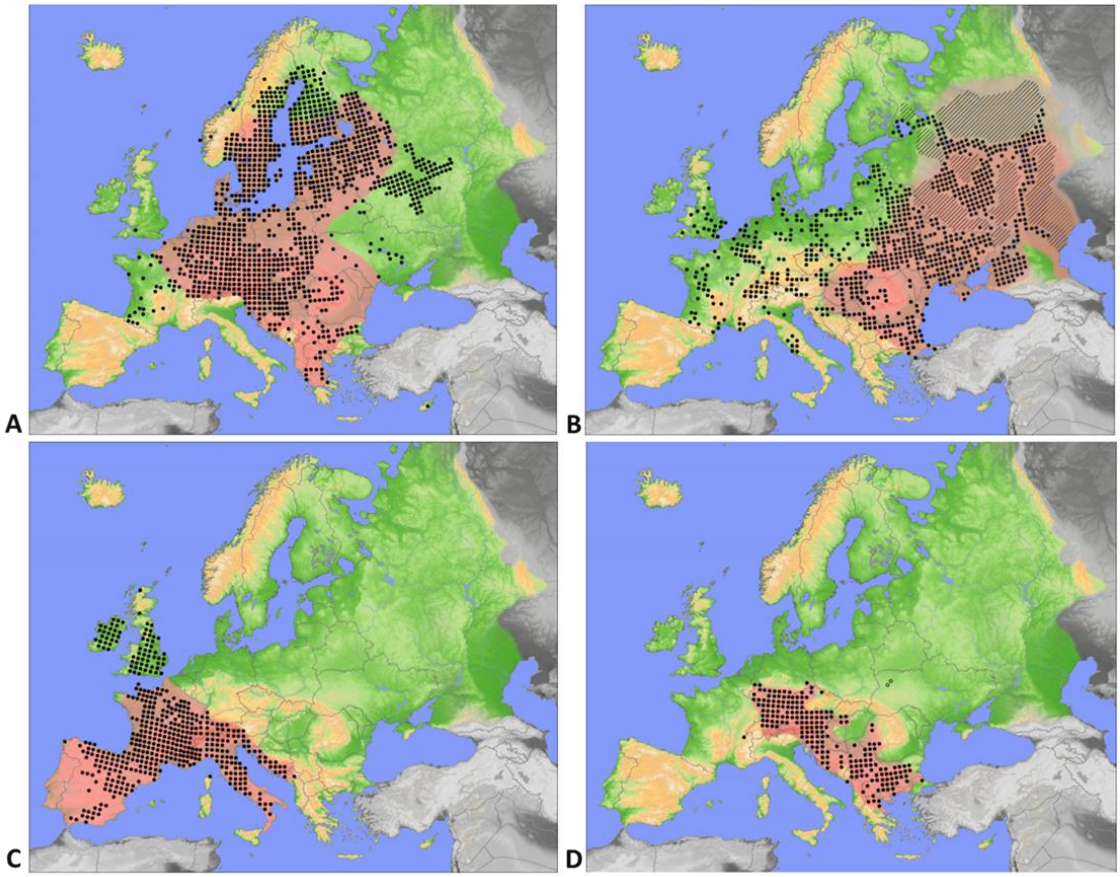


Figure 1.5 Geographical distribution of the four European crayfish affected by the crayfish plague. **A**, *Astacus astacus*; **B**, *Astacus leptodactylus*; **C**, *Austropotamobius pallipes*; **D**, *Austropotamobius torrentium*. Presumed native range highlighted in red; balck circles indicate most recent reported populations. The thick-clawed crayfish (*Astacus pachypus*) has not been included in the figure as no crayfish mortalities caused by the crayfish plague have been recorder to date. Figure modified from Kouba *et al*. (2014).

Although crayfish stock declines and mortalities can be caused by habitat deterioration, poor water quality and pollution, the introduction of North American crayfish species in Europe for recreational and farming purpose played an important role in the extensive damage to European crayfish stocks. Besides being more aggressive and fast growing in comparison to the European species, North American species carry *A. astaci*, the oomycete causative agent of crayfish plague, fatal disease for the European crayfish species (Alderman, 1996; Edsman *et al*., 2010; OIE, 2016a; Söderhäll and Cerenius, 1999). While there are no recorded mortalities caused by the crayfish plague in the thick-clawed crayfish (*Astacus pachypus*), the other four European crayfish species have all been affected by this disease, which caused significant declines in population numbers. Three out of five of these species are now listed in the IUCN Red List of Threatened Species (http://www.iucnredlist.org; Souty-Grosset *et al*., 2006; Holdich *et al*., 2009).

The introduction in Europe of North American crayfish carrying *A. astaci*, thus caused the decline of European crayfish and extensive damage to both wild and cultured European crayfish stocks. Although crayfish mortalities associated to *A. astaci* have been reported throughout Europe, well documented crayfish catchment declines have been recorded only in Turkey, Sweden, and Finland. In particular, Turkey's indigenous crayfish *Astacus leptodactylus* (narrow-clawed crayfish) harvests declined from 7,000 metric tonnes in 1984 to 300 metric tonnes in the early 90s (Ackefors, 1999; Harlioğlu and Harlioğlu, 2006); Finland catchments of native noble crayfish (*Astacus astacus*) exceeded the 16 million at the beginning of the 20th century, but dropped to 2 million in the 21st century (Jussila and Mannonen, 2004); in Sweden, Edsman (2004) estimated that only 5% of the original noble crayfish (*A. astacus*) populations were left in the year 2000, wiped out in only 100 years by the crayfish plague.

The decline of European crayfish not only had an impact on human's economic activities, but also on the ecosystem. Crayfish represent an important component of the freshwater fauna and, more importantly, can control the biodiversity (flora and fauna) of the surrounding environment by their feeding habits (carnivorous, herbivorous, detritivores or omnivores) (Reynolds *et al*., 2013). For example, the native European crayfish occupies the role of top predator in some Swedish lakes. Thus, the decline (or extinction) of this top

predator triggered by *A. astaci* can alter the proportion of macrophytes and other invertebrates in the aquatic environment, causing imbalances between species and reduce the biodiversity of the ecosystem (Reynolds and Souty-Grosset, 2012).

1.2.3 Infection and symptoms of the crayfish plague

As previously mentioned in section 1.1.1, primary cysts of *A. astaci* release secondary zoospores, which start to swim and locate a host to infect. The infection process starts when the swimming zoospore attaches and encysts onto the host surface and begins to germinate (Cerenius and Söderhäll, 1984). Germination is mediated by a germination peg that penetrates the cuticle of the host and forms a new hypha (Nyhlén and Unestam, 1975). As no spines or mucous substances have been identified on *A. astaci* secondary zoospores surface that can aid the interaction with the host, the most successful infections are located on the soft abdominal cuticle and joints, or on damaged cuticles of the crayfish, where the zoospores are more likely to be trapped by mechanical forces (Nyhlén and Unestam, 1975; Unestam and Weiss, 1970). Once an infection has established, in both European crayfish species (*e.g. A. astacus* – noble crayfish) and North American crayfish species (*e.g. Pacifastacus leniusculus* – signal crayfish), *A. astaci* penetrates the cuticle and starts to produce new hyphae while the host haemocytes enclose the infection by aggregating and surrounding the cysts and hyphae (Unestam and Weiss, 1970). In both European and North American crayfish species, deposits of melanin produced by the haemocytes in the cuticle surrounding the pathogens have been observed (Figure 1.6). This melanisation is the host defence to enclose and limit any mechanical damage or damages caused by an infectious agent, including *A. astaci* (Cerenius *et al.*, 2003). Melanisation stimulated by *A. astaci* is more abundant and active in North American crayfish species than European species (Cerenius *et al.*, 2008; Nyhlén and Unestam, 1975). By surrounding *A. astaci* in the melanised cuticle, North American crayfish species can carry the live parasites as a chronic infection. On the contrary, European crayfish cannot control the infection efficiently, resulting in a deadly and fatal condition that usually reaches 100 % mortality (Alderman *et al.*, 1987; Nyhlén

and Unestam, 1975; Unestam, 1972). Clinical signs of infection in European crayfish are loss of co-ordination, reduced escape reflexes, whitening of the muscles near the infection point and melanisation of the cuticle (Alderman *et al.*, 1987). Even if North American crayfish can control and manage *A. astaci* infection in normal conditions, crayfish plague outbreaks have been reported in co-infection or when the animals are subjected to stressful conditions, indicating that North American crayfish are not only asymptomatic carriers but can also develop disease when their immune system is challenged or suppressed (Aydin *et al.*, 2014; Persson and Söderhäll, 1983; Thörnqvist and Söderhäll, 1993).

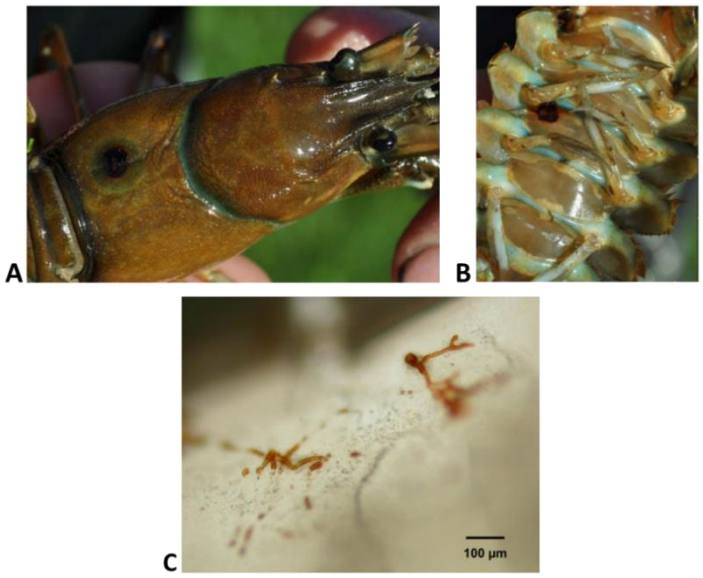


Figure 1.6 External signs of crayfish plague. **A** and **B**, melanised cuticles of *Pacifastacus* leniusculus (signal crayfish) caused by *A. astaci* infection, from Vrålstad *et al*. (2011). **C**, melanised *A. astaci* hyphae in the cuticle of *Orconectes limosus* (spiny-cheek crayfish), from Kozubíková *et al.* (2011).

1.2.4 Diagnosis of crayfish plague

Due to the unsuccessful isolation of the aetiological agent of crayfish plague from the first European freshwater crayfish mass mortalities, *A. astaci* was not confirmed as the causative agent of crayfish plague until 1934 when Nybelin (1934) and Rennerfelt (1936) successfully isolated the parasite, 70 years after its first appearance in Europe. Even after the improvement of isolation techniques, in more recent outbreaks diagnosed as crayfish plague the isolation of the pathogen failed (Diéguez-Uribeondo *et al.*, 1997; Edgerton *et al.*, 2004; Taugbøl *et al.*, 1993). A recent investigation on epizootic events in European

freshwater crayfish showed that in several mortalities other potential crayfish pathogens were present, *e.g.* fungi (Söderhäll *et al.*, 1993), viruses (Edgerton *et al.*, 1996; Paasonen *et al.*, 1999), other oomycetes (Diéguez-Uribeondo *et al.*, 1997; Söderhäll *et al.*, 1991), bacteria and protists (Edgerton *et al.*, 2004). Thus, mortalities without proper identification of the parasite cannot be directly linked to crayfish plague and *A. astaci*. Until 2006, the isolation of *A. astaci* and transmission trials were necessary to confirm crayfish plague diagnoses, as there are no morphological features to distinguish *A. astaci* from other saprotrophic/opportunistic *Aphanomyces* species (OIE, 2016a; Vrålstad *et al.*, 2014). On the other hand, the development of PCR and qPCR (Oidtmann *et al.*, 2006; Vrålstad *et al.*, 2009) targeting *A. astaci* ITS regions have improved the crayfish plague diagnostic tool set and these techniques now allow a reliable discrimination not only of *A. astaci* pure cultures, but also of fresh and preserved crayfish samples infected with *A. astaci*. These methods have now been accepted as diagnostic standards (Grandjean *et al.*, 2014; Kozubíková *et al.*, 2011; OIE, 2016a; Tuffs and Oidtmann, 2011; Vrålstad *et al.*, 2011, 2014) and have been used to monitor indigenous and non-indigenous crayfish species in more recent years (Caprioli *et al.*, 2013; Filipová *et al.*, 2013; Hsieh *et al.*, 2016; Kušar *et al.*, 2013; Maguire *et al.*, 2016; Marino *et al.*, 2014; Pârvulescu *et al.*, 2012; Pretto *et al.*, 2014; Viljamaa-Dirks *et al.*, 2011).

1.2.5 *Aphanomyces astaci* genotypes

Wilhelm Johannsen, a Dutch genetist and botanist, firstly introduced the term *genotype* in 1911, meaning the total sum of all genes in a gamete or zygote (Johannsen, 1911). Since then, the concept of genotype has been adapted to refer the DNA sequence of a cell (or individual), which determines the phenotype of that cell (or individual), including gene variants (alleles) carried by the cell (or individual) (Taylor and Lewontin, 2017). By this last definition, a genotype can be considered as the amount of differences between individuals belonging to the same species, differences which can be identified by genotyping techniques (Hehenberger, 2015). Thanks to the application of the genotyping technique RAPD-PCR on *A. astaci* pure cultures, genetic differences between isolates have been identified, allowing the distinction of a

total of five genotypes (Table 1.1): genotype A (or As), isolated from *A. astacus* and *A. leptodactylus* which includes *A. astaci* isolates related to the first introductions of this parasite in Europe; genotype B (or PsI), firstly isolated from *P. leniusculus* and from European indigenous freshwater crayfish after the introduction of *P. leniusculus* in Europe; genotype C (or PsII), based on one single isolate from *P. leniusculus* native in Canada and imported into Sweden; genotype D (or Pc), which includes isolates from *Procambarus clarkii* and more recently from *Austropotamobius pallipes*; genotype E (or Or), isolated from *Orconectes limosus* (Diéguez-Uribeondo *et al.*, 1995; Huang *et al.*, 1994; Kozubíková, *et al.*, 2011; Lilley *et al.*, 1997a; Rezinciuc *et al.*, 2014). The nomenclature of the genotypes used in the present study (A to E) follows the original genotype division by Huang *et al.* (1994). The alternative nomenclature of *A. astaci* genotypes (As to Or) is based on the initials of the North American host from which the genotypes have been originally isolated, except for genotype A (or As), which it has been firstly isolated from *A. astacus* while the North American crayfish host is still unknown.

Table 1.1 *Aphanomyces astaci* genotypes. Division of genotypes is based on RAPD-PCR analysis. Original North American crayfish host for genotype from where the culture was firstly isolated is indicated.

Genotype	**Original North American host**
A (As)	Unknown (Huang *et al*., 1994)
B (PsI)	Signal crayfish, *Pacifastacus leniusculus* (Huang *et al*., 1994)
C (PsII)	Signal crayfish, *Pacifastacus leniusculus* (Huang *et al*., 1994)
D (Pc)	Red swamp crayfish, *Procambarus clarkii* (Diéguez-Uribeondo *et al*., 1995)
E (Or)	Eastern Crayfish, *Orconectes limosus* (Kozubíková *et al.*, 2011)

Epidemiological studies of crayfish plague dependent on the distinction of *A. astaci* genotypes proved to be useful tools to better understand the history and spread of this disease in Europe (Kozubíková *et al.*, 2008; Lilley *et al.*, 1997a; Rezinciuc *et al.*, 2014; Viljamaa-Dirks *et al.*, 2013). While these studies were based on *A. astaci* pure cultures, the application of microsatellite typing allow the discrimination of *A. astaci* genotypes directily from infected crayfish material, enabling new epidemiological studies on North American carrier samples (Filipová *et al.*, 2013; Grandjean *et al.*, 2014; Maguire *et al.*, 2016; Vrålstad *et al.*, 2014). Most importantly, infection studies on both European and North American crayfish indicate differences in virulence within genotypes (A and B) and within different isolates belonging to the same genotype (A) (Aydin

et al., 2014; Becking *et al.*, 2015; Jussila *et al.*, 2013; Makkonen *et al.*, 2014; Makkonen *et al.*, 2012; Viljamaa-Dirks *et al.*, 2016).

Epidemiological studies and virulence studies combined, showed an increased resistance of some populations of *A. astacus* against the “old” genotype A, thought to be first one introduced into Europe. This could signify the presence of possible co-evolution between European crayfish and *A. astaci* genotype A in the last 150 years, suggesting the presence of *A. astacus* populations carrying the disease in a chronic manner (Makkonen *et al.*, 2014; Makkonen, *et al.*, 2012; Viljamaa-Dirks *et al.*, 2013). For example, recent epidemiological surveys on crayfish plague in Finland demonstrate that populations weakened by previous crayfish plague epizootic events can carry genotype A as a form of persistent chronic infection. Some of these populations subsequently faded due to the arrival of North American crayfish carrying genotype B (Viljamaa-Dirks *et al.*, 2013). This hypothesis is supported by the studies of Makkonen *et al.* (2012) indicating an increased *A. astacus* resistance against genotype A due to co-evolution. Moreover, in Finland some repopulation programmes failed due to epizootic co-infections caused by both A and B genotypes (Viljamaa-Dirks *et al.*, 2013). These virulence differences could be due to genetic variants that separate *A. astaci* genotypes. Studying this disease using molecular epidemiology is still important to understand the spread of crayfish plague in Europe, to prevent new epizootic events and to track possible source of infections. Reliable genotyping methods to characterise *A. astaci* are still needed as the current genotyping techniques present some deficiencies.

A brief description of the current genotyping tools is given below, while the reader is directed to sections 3.3 and 3.4 for the new genotyping techniques developed in the present study.

1.3 Genotyping techniques for *Aphanomyces astaci*

The PCR based and sequencing methods for the diagnosis of crayfish plague do not allow the discrimination of genotypes (Makkonen *et al.*, 2011) and different fingerprinting tools to detect DNA polymorphisms have been developed. Four tools (RAPD-PCR, amplified fragment length polymorphism (AFLP), chitinase genes, and microsatellites) are currently used to genotype *A. astaci*, however, they all have deficiencies. The new methods developed in the present study, which are described in sections 3.3 and 3.4, overcome these deficiencies.

1.3.1 RAPD-PCR

The first and most used genotyping tool for *A. astaci* pure cultures is the RAPD-PCR developed by Huang *et al.* (1994). This technique randomly amplifies DNA regions that are not necessarily known to the user (Ali *et al.*, 2004), and until recently, it has been the fundamental classifying tool for all known *A. astaci* genotypes. While this technique can detect large numbers of DNA polymorphisms allowing the comparison of genetically closely related individuals belonging to the same species, it presents some downsides that can influence its reproducibility and repeatability, *e.g.* the standardisation of the protocol, the low stringency of homology of the primers used, the concentration of DNA template and the purity of the isolate (Jones *et al.*, 1997; Penner *et al.*, 1993; Skroch and Nienhuis, 1995). For a detailed discussion of the results of this method applied to the *A. astaci* cultures used in this study, the reader is referred to section 3.1.

1.3.2 AFLP

The AFLP for *A. astaci* has been developed by Rezinciuc *et al.* (2014). This technique involves the amplification of the DNA with a set of generic primers and, as the RAPD-PCR, it does not require prior knowledge of the targeted DNA sequence (Vos *et al.*, 1995). The method has been successfully applied to *A. astaci* pure cultures isolated during crayfish plague outbreaks in Spain (Rezinciuc *et al.*, 2014). The authors noticed that, while supporting the RAPD-

PCR groupings, the technique could identify minor variations within the genotypes tested (A, B, C, and D). While this fingerprinting method is more robust and stringent than the RAPD-PCR, it can be applied only to pure cultures that are, however, difficult to obtain and maintain under laboratory conditions.

1.3.3 Chitinase genes

Another method used to detect and evaluate genetic diversity between *A. astaci* genotypes is the amplification and sequence analysis of chitinase genes (Makkonen *et al.*, 2012). These genes are of particular interest in *A. astaci*, as they are an indication of virulence and might be expressed during infection (Hochwimmer *et al.*, 2009). In contrast with the RAPD-PCR and AFLP, the authors could successfully amplify and analyse the targeted genes from infected material (bypassing therefore the pure culture step). They usefully identified sequence differences between genotypes and between isolates belonging to the same genotype, emphasising the evolutionary importance of these genes. Unfortunately, as for the AFLP, only four (A, B, C, and D) of five genotypes were analysed and the variations observed were not able to distinguish all four (Makkonen *et al.,* 2012).

1.3.4 Microsatellites

Microsatellites are regions of DNA consisting of 1-10 nucleotides (repeat unit) repeated 5-50 times (Gemavel *et al.*, 2012). These regions are widely distributed in the genome of both eukaryotes and prokaryotes organisms and are of special interest in populations studies as they present a higher mutation rate in comparison to point mutations (Gemavel *et al.*, 2012; Pérez-Jiménez *et al*., 2013; Phumichai *et al*., 2015). The most common mutation in microsatellites loci is due to polymerases errors (*e.g.* strand-slippage) which causes the addition or the deletion of a repeat unit during replication, producing a different number of repeat units in individuals belonging to the same species and genotype (Carneiro Vieira *et al.,* 2016). Microsatellites have been successfully applied in plant and fish population genetics and genotyping but also in human pathology (Abdul-Muneer, 2014; Carneiro Vieira *et al.,* 2016; Gras *et al*., 2000)

and more recently they have also been applied to the oomycete *A. astaci* by Grandjean *et al.* (2014) in order to develop a method to genotype this parasite. This method bypasses the pure culture step, allowing genotyping *A. astaci* directly from infected host tissues, preserved or fresh, both from European and North American crayfish. However, the authors recommend conducting other tests to confirm the presence of *A. astaci* in the samples analysed, as the presence of unknown saprotrophic oomycetes species can lead to difficult interpretations of results. As for AFLP, microsatellites typing supports the RAPD-PCR grouping but various inconsistencies were observed for genotype A, B, and D (James *et al.*, 2017; Maguire *et al.*, 2016; Mrugała *et al.*, 2016; Vrålstad *et al.*, 2014), discrepancies that could indicate the presence of an unknown *A. astaci* genotype, the presence of a variation in the genotype, or a deficiency of the tool. This fingerprinting tool comes with some limitations, such as the presence of null or partial null alleles due to primers misaligning or the preferred amplification of some alleles rather than others (Dakin and Avise, 2004; Jarne and Lagoda, 1996). For a comparison between the microsatellite method and the new genotyping methods developed in this study, the reader is referred to section 3.4.8.

1.4 Next generation sequencing

DNA sequencing started in the 1970s, when Sanger *et al.* (1973) developed the first polymerase-based sequencing method. This first method and its implementations, enabled the whole genome sequencing (WGS) of various organisms, such as viruses (*e.g.* bacteriophages), prokaryotes (*e.g.* archaebacteria and bacteria), unicellular eukaryotes (*e.g.* yeasts), multicellular eukaryotes (*e.g.* round worms, plants, insects, and humans) (de Magalhães *et al.*, 2010; IHGSC, 2004; Metzker, 2005; Studholme *et al.*, 2011). Despite improvements and achievements, Sanger sequencing methods have now been largely replaced by next generation sequencing (NGS) methods, which enabled the fast and cost-effective sequencing of organisms' genomes by producing large quantities of genomic data in a shorter time (Sabat *et al.*, 2013). NGS has transformed and empowered biological and medical research and has been successfully applied to various fields of biology, pathology, epidemiology, microbial evolution, outbreaks characterisation, host-pathogen interaction, forensics, applied medicine and populations genetics (Grad *et al.*, 2012; Gudbjartsson *et al.*, 2015; Hao *et al.*, 2012; Jiang *et al.*, 2013; Liti *et al.*, 2009; Metzker, 2005; Sabat *et al.*, 2013; Studholme *et al.*, 2011).

NGS has had an important impact on genetic studies by offering a cheaper way to discover genome-wide variations, with the resolution of single nucleotides (Sabat *et al.*, 2013). This information has been used, for example, to identify variations in cancerous cells, to detect new genetic markers, to identify differences between closely related organisms and to develop new genotyping techniques (Schadt *et al.*, 2010; Studholme *et al.*, 2011). Various studies successfully used WGS data to generate lists of single nucleotide variants (SNVs) to separate and characterise outbreaks of closely related isolates, such as the methicillin-resistant *Staphylococcus aureus* (Kong *et al.*, 2016; Köser *et al.*, 2012), or the discovery of unique genome regions and virulence factors in *Escherichia coli* outbreaks that helped the identification and design of strain-specific assays for the characterisation of the pathogen (Rasko *et al.*, 2011). WGS has been also used to develop or improve genotyping techniques in *Mycobacterium tuberculosis*, pea, and rice (Boutet *et al.*, 2016; Kim *et al.*, 2016; Roetzer *et al.*, 2013), concepts followed in this study to develop new genotyping assays able to distinguish all known *A. astaci* genotypes (sections 3.3 and 3.4).

One of the suppliers of NGS is Illumina, platform which has been employed in this study. With this platform, the DNA template to be used for sequencing is firstly randomly fragmented, ligated with adapters, amplified by PCR and gel purified, in a process called library preparation. The library is then loaded into the machine flow cell, where the fragments are immobilised to a solid phase to generate the clusters. Then, the templates are sequenced by imaging dye-labelled nucleotides ligated to each template and washed away in consecutive cycles (Metzker, 2010). Once the reads are generated, they are either aligned to a reference sequence (if available) or *de novo* assembled, depending on the purpose of the research. For example, if an organism's genome reference sequence is already available and the intent of the research is to identify genetic variation between closely related strains, simply aligning the newly generated reads would be a preliminary sufficient approach to compare the datasets. In the case where a full reference genome is not available, the researcher will need to *de novo* assemble those reads, a process which involves bioinformatically joining the reads into contiguous sequences called contigs and scaffolds. Numerous assemblers and algorithms are available to assemble NGS reads. Here I mention the de Bruijn graph algorithm, which has been used to assemble the datasets generated in this study. In the de Bruijn approach, the assembler firstly builds a graph of substrings extracted from the reads of a given *k* length (*k*-mer). Once the graph is set, the reads paths are traced back along the graph, while the assembler uses the reads information to resolve the structure and constructing contigs (Nagarajan and Pop, 2013; Nielsen *et al.*, 2011). Despite the ongoing development of algorithms and assemblers driven by NGS technologies, the process of assembling a new genome is not error free and the result is a hypothesis or an assumption of the sequence of the target DNA (Nagarajan and Pop, 2013; Studholme, 2016). Therefore, it is important to assess the quality of the reads before attempting to assemble a genome and to assess the quality of newly assembled genome.

The reader is redirected to section 3.6 for information on the WGS, assembled genomes and investigation on *A. astaci* genotypes differences with a bioinformatic approach completed in this study.

1.5 Aims of the PhD project

Aphanomyces astaci is the causative agent of crayfish plague, a disease listed by the World Organisation for Animal Health (OIE) and fatal for European crayfish species. It was first introduced into Italy from the US in the late 19th century and rapidly spread throughout Europe causing the decline of native crayfish. It currently threatens to wipe out the UK native white-clawed crayfish (*Austropotamobius pallipes*). The application of RAPD-PCR to pure cultures has genetically characterized five genotypes. This distinction proved to be a useful tool for epidemiological studies aimed at understanding the history and spread of the disease in Europe.

While the available genotyping tools can distinguish all known *A. astaci* genotypes, there are some practical drawbacks to genotype by the current genotyping methods. The present study hypothesises that the genetic diversity between genotypes can be further studied and new reliable genetic markers can be developed and used for epidemiological studies. Ideally, these markers should bypass the culture methods, differentiate between species of *Aphanomyces* and detect different genotypes and strains that can be present at the same time in the same sample. Therefore, WGS was performed on *A. astaci* cultures held in the Centre of Environment, Fisheries and Aquaculture Science (Cefas) Oomycetes Culture collection (OCC) to investigate the genome of *A. astaci* isolates belonging to different genotypes and identify genetic differences that might serve as new phylogenetic markers to distinguish genotypes. DNA single nucleotide variants (SNVs) and genotype-unique genomic regions were catalogued. These were exploited as phylogenetic markers and were tested both on pure cultures and historical samples derived from outbreaks and carrier crayfish available in our laboratories. These genotyping methods represent new diagnostic tools aiding the detection and prevention of crayfish plague.

2. Materials and Methods

2.1 Materials

All chemicals, unless stated otherwise, were obtained from Sigma-Aldrich. All solutions were made using distilled water. All media and buffer used for microbiology were sterilised by autoclaving at 121 °C for 10 minutes or by filter sterilisation using Sartorius Minisart filters 0.2 µm. pH was adjusted with 1 M sodium hydroxide or 1 M hydrogen chloride, unless stated otherwise.

2.1.1 Cultured *Aphanomyces* isolates

Aphanomyces isolates (Table 2.1) used in this study were provided by Cefas OCC.

Table 2.1 List of isolates provided by Cefas OCC and tested in this study. Host and geographical area of isolation are indicated when known. Note: *A. invadans*-like NJM9510 has been submitted to the American Type Culture Collection (ATCC number 62427) by Dykstra *et al.* (1986). - Information not know. * Genotypes are here defined as per the Cefas OCC record system.

Species	**Isolate (Genotype)***	**Host of isolation Scientific and common name**	**Country of isolation**
A. astaci	Da (A)	-	-
	SV (A)	*Astacus astacus* - noble crayfish	Sweden
	457 (B)	*Austropotamobius pallipes* - white-clawed crayfish	United Kingdom
	197901(B)	*A. pallipes* - white-clawed crayfish	United Kingdom
	D2 (B)	*A. astacus* - noble crayfish	Germany
	SA (B)	*A. astacus* - noble crayfish	Germany
	Si (B)	-	-
	YX (B)	*A. astacus* - noble crayfish	Sweden
	KV (C)	*Pacifastacus leniusculus* - signal crayfish	Finland
	Pc (D)	*Procambarus clarkii* - red swamp crawfish	Spain
	KB13 (E)	*Orconectes limosus* - spinycheek crayfish	Czech Republic
A. frigidophilus	AP5	*P. leniusculus* - signal crayfish	origin of crayfish: Sweden; country of isolation: Germany
	RP1	*Oncorhynchus mykiss* - rainbow trout	United Kingdom
	RP2	*O. mykiss* - rainbow trout	United Kingdom
A. invadans	GWR	-	United States of America
	NJM0002	-	-
	NJM8997	*Plecoglossus altivelis* - ayu	Japan
	NJM9030	*P. altivelis* - ayu	Japan
	NJM9701	*P. altivelis* - ayu	Japan
"*A. invadans*-like"	NJM9510	*Brevoortia tyrannus* - Atlantic menhaden	United States of America

2.1.2 Primers

Primers used in this study are described in Table 2.2. All primers were designed by Diana Minardi unless otherwise stated. Genotyping primers for *A. astaci* isolates were chosen by visualisation of sequence alignments around fragments

of interest in the Integrative Genomics Viewer (IGV) (Robinson *et al.*, 2011; Thorvaldsdóttir *et al.*, 2013). Prior to synthesis, primers were tested for secondary structures formation using the program Oligo Analysis Tool (Eurofins Genomics). Optimal aligning temperatures for each primer pairs were checked by gradient PCR. Eurofins Genomics provided primer synthesis.

Table 2.2 List of primers used in this study. Assigned names, primer sequences, references, and purposes of use indicated. COI, cytochrome oxidase subunit I; ITS, internal transcribed spacer (partial 18 S rDNA gene, ITS1 region, 5.8 S rDNA gene, ITS2 region and partial 28 S rDNA gene); LSU, large subunit ribosomal DNA (partial 18 S rDNA gene, ITS1 region, 5.8 S rDNA gene, ITS2 region and partial 28 S rDNA gene); SNV, single nucleotide variant; mtDNA, mitochondrial DNA.

Primer name	**Sequence (5'-3')**	**Reference**	**Purpose**
B01	GTTTCGCTCC	Huang *et al.* (1994)	RAPD-PCR
ITS1	TCCGTAGGTGAACCTGCGG	White *et al.* (1990)	Amplification of ITS, forward primer
ITS5	GGAAGTAAAAGTCGTAACAAGG	White *et al.* (1990)	Amplification of ITS, forward primer
ITS4	TCCTCCGCTTATTGATATGC	White *et al.* (1990)	Amplification of ITS and sequencing, reverse primer
Oom-COI-Lev-up	TCAWCWMGATGGCTTTTTTCAAC	Bala *et al.* (2010)	Amplification of COI, forward primer
Fm85mod	RRHWACKTGACTDATRATACCAAA	Bala *et al.* (2010)	Amplification of COI, reverse primer
UNup18S42	CGTAACAAGGTTTCCGTAGGTGAAC	Bakkeren *et al.* (2000)	Amplification of LSU, forward primer
UN-lo28S1220	GTTGTTACACACTCCTTAGCGGAT	Bala *et al.* (2010)	Amplification of LSU, reverse primer
IF	TTGAAGGCAGAATGCGGAGT	This study	Sequencing, internal forward primer
95	TGCATTTGTGTTGACCGTGG	This study	Sequencing, internal forward primer
FRIG	TCGCAAAATGCGGAGTGAGA	This study	Sequencing, internal forward primer
IR	CATTTCGCCAGAGTCCCGAA	This study	Sequencing, internal reverse primer
pJET1.2for	CGACTCACTATAGGGAGAGCGGC	Eurofins Genomics	Sequencing forward primer of inserts in pJET1.2/blunt Cloning Vector (Thermo Scientific)
pJET1.2rev	AAGAACATCGATTTTCCATGGCAG	Eurofins Genomics	Sequencing reverse primer of inserts in pJET1.2/blunt Cloning Vector (Thermo Scientific)
M13 uni (-43)	AGGGTTTTCCCAGTCACGACGTT	Eurofins Genomics	Sequencing forward primer of inserts in pGEM-T-Easy vector (Promega)
M13 rev (-29)	CAGGAAACAGCTATGACC	Eurofins Genomics	Sequencing forward primer of inserts in pGEM-T-Easy vector (Promega)
A_HhaI_F	CAACGACGGCCTCTTGTACCATTC	This study	Amplification of nuclear DNA with SNV in genotype A and sequencing, forward primer
A_HhaI_R	GCCAGCTAGTCAACAGAGATAACG	This study	Amplification of nuclear DNA with SNV in genotype A and sequencing, reverse primer
B_HhaI_F	GCATGAGTCCAGATTCGAGGTG	This study	Amplification of nuclear DNA with SNV in genotype B and sequencing, forward primer
B_HhaI_R	GTTACCAGTACACTCGGAGAGAGC	This study	Amplification of nuclear DNA with SNV in genotype B and sequencing, reverse primer
C_HhaI_F	CCATCCGTCAAAAGCTGCAATCAG	This study	Amplification of nuclear DNA with SNV in genotype C and sequencing, forward primer
C_HhaI_R	CGCCGACTTCATTTGTTTACCGGT	This study	Amplification of nuclear DNA with SNV in genotype C and sequencing, reverse primer
D_HhaI_F	CCATCGACAAGTTGTTTGCCTTG	This study	Amplification of nuclear DNA with SNV in genotype D and sequencing, forward primer
D_HhaI_R	GTGAGCACGGCATTGTAAATTTGC	This study	Amplification of nuclear DNA with SNV in genotype D and sequencing, reverse primer
E_HhaI_F	GGAGCTACGCAAGTGCTGCAACCA	This study	Amplification of nuclear DNA with SNV in genotype E and sequencing, forward primer
E_HhaI_R	CATGATAAATCGCTGGTACTCTGG	This study	Amplification of nuclear DNA with SNV in genotype E and sequencing, reverse primer
A_MITO_F	CCAAATTCTCCTTTAGGCGCTTC	This study	Amplification of mtDNA with SNV in genotype A and sequencing, forward primer
A_MITO_R	CAGGAGCTCGTATGCATTCAAGTT	This study	Amplification of mtDNA with SNV in genotype A and sequencing, reverse primer
B_MITO_F	GGAGCATGTGGTTTAATTCGACAA	This study	Amplification of mtDNA with SNV in genotype B and sequencing, forward primer
B_MITO_R	CAAAGTCTAGTAACATATTCACCGC	This study	Amplification of mtDNA with SNV in genotype B and sequencing, reverse primer
B_MITO_N1S	TCATGGCCCTTATGGG	This study	Semi-nested forward primer for amplification of mtDNA with SNV in genotype B
B_MITO_N2S	TCATGGCCCTTATGGA	This study	Semi-nested forward primer for amplification of mtDNA without SNV in genotype B
C_MITO_F	CCTTTACAGTACTTGTTCACTATCGG	This study	Amplification of mtDNA with SNV in genotype C and sequencing, forward primer

C_MITO_R	GCATAGAGGGGATGCCTAGGC	This study	Amplification of mtDNA with SNV in genotype C and sequencing, reverse primer
D_MITO_F	GCTCCTGGTATTATGCCTAGACAA	This study	Amplification of mtDNA with SNV in genotype D and sequencing, forward primer
D_MITO_R	GGATAAGCTTCTCTACCTGGAGG	This study	Amplification of mtDNA with SNV in genotype D and sequencing, reverse primer
E_MITO_F	CGTGTATTTGAGACAAATAAGCCA	This study	Amplification of mtDNA with SNV in genotype E and sequencing, forward primer
E_MITO_R	GGAACTTACCCGACAAGGAATTTC	This study	Amplification of mtDNA with SNV in genotype E and sequencing, reverse primer
A_unique_F	GCAACTTCCACGTAGTTACAAATC	This study	Amplification of genotype A specific DNA region and sequencing, forward primer
A_unique_R	GCTGCTGGCTACTTCTCAGTGTT	This study	Amplification of genotype A specific DNA region and sequencing, reverse primer
B_unique_F	CGAAAAGCTCGAGAACGCAGAGC	This study	Amplification of genotype B specific DNA region and sequencing, forward primer
B_unique_R	CGTTCCTCTTCAGTGTAGCGCTC	This study	Amplification of genotype B specific DNA region and sequencing, reverse primer
C_unique_F	CCAGCAACATACCAGTTGCGAACG	This study	Amplification of genotype C specific DNA region and sequencing, forward primer
C_unique_R	GCACAATTTGATCCCTCTTTTCTG	This study	Amplification of genotype C specific DNA region and sequencing, reverse primer
D_unique_F	GCAAGAAGTAAAGGATATTTATTC	This study	Amplification of genotype D specific DNA region and sequencing, forward primer
D_unique_R	GCTTATAGTATCTACATTTTCCGC	This study	Amplification of genotype D specific DNA region and sequencing, reverse primer
E_unique_F	GCAGTTGAGATCCATCTTCATCAT	This study	Amplification of genotype E specific DNA region and sequencing, forward primer
E_unique_R	CGAAACTACGTCCTAATCAACACA	This study	Amplification of genotype E specific DNA region and sequencing, reverse primer

2.1.3 Cloning PCR products in *Escherichia coli*

The PCR-amplified fragments for the internal transcribed spacer of the ribosomal DNA (ITS), the D1-D2 regions of the large subunit ribosomal DNA (LSU) and the cytochrome oxidase subunit I (COI) and RAPD-PCR products were cloned in pJET1.2/blunt Cloning Vector (Thermo Scientific) or in pGEM-T-Easy vector (Promega) and transformed in *E. coli* DH5α™ competent cells (Invitrogen).

2.1.4 Media and solutions for *Aphanomyces* cultures

The following media were used in this study as routine sub culturing media for *Aphanomyces* isolates.

Glucose-Peptone (GP) broth

From Lilley *et al.* (1998)

3 g glucose anhydrous

1 g peptone bacteriological (Oxoid)

0.128 g magnesium sulphate ($MgSO_4.7H_20$)

0.014 g potassium dihydrogen phosphate (KH_2PO_4)

0.029 g calcium chloride ($CaCl_2.2H_20$)

0.0024 g iron chloride ($FeCl_3.6H_2O$)

0.0018 g manganese chloride ($MnCl_2.4H_20$)

0.0039 g copper sulphate ($CuSO_4.5H_20$)

0.0004 g zinc sulphate ($ZnSO_4.7H_2O$)

distilled water up to 1000 ml

pH 7

The broth was stored at room temperature after autoclaving, until required.

GP agar plates

3 g glucose anhydrous

From Lilley *et al.* (1998)	12 g agar technical (Oxoid)
	1 g peptone bacteriological (Oxoid)
	0.128 g magnesium sulphate ($MgSO_4.7H_20$)
	0.014 g potassium dihydrogen phosphate (KH_2PO_4)
	0.029 g calcium chloride ($CaCl_2.2H_20$)
	0.0024 g iron chloride ($FeCl_3.6H_2O$)
	0.0018 g manganese chloride ($MnCl_2.4H_20$)
	0.0039 g copper sulphate ($CuSO_4.5H_20$)
	0.0004 g zinc sulphate ($ZnSO_4.7H_2O$)
	distilled water up to 1000 ml
	pH 7

The agar was spread into sterile 12 cm plates in a flow cabinet after autoclaving and left to solidify. Plates were stored at 4 °C until required.

GP broth and agar are two of the media used by the scientific communities working with *Aphanomyces* species. These media are used especially for *A. invadans*, freshwater parasite of fish (OIE, 2016b) and the recipe used in this study follows Lilley *et al.* (1998) adaptation of Willoughby and Roberts (1994) recipe. Interestingly, this media does not contain any sodium chloride, which is used by the cells to maintain electrolyte and osmotic balance. *Aphanomyces* cultures are notoriously difficult to obtain and maintain alive for a long time in laboratory conditions (Oidtmann *et al.*, 1999) and this could be due to suboptimal growing media. In this study, the cultures provided by Cefas OCC were grown and subcultured in GP media only for the time necessary to obtain enough high molecular DNA subsequently used for whole genome sequencing and not for a long-term conservation plan. Therefore, more suitable long-term growing media were not investigated.

2.1.5 Media and solutions for bacterial cultures

The following media and solutions were used in this study.

Luria-Bertani (LB) broth	25 g premixed powder (Melford) distilled water up to 1000 ml
LB plates	25 g premixed powder (Melford) 8 g agar (Melford) distilled water up to 1000 ml
Super optimal broth (SOB)	20 g tryptone (Melford) 5 g yeast extract 0.5 g sodium chloride (NaCl) 2.4 g magnesium sulphate anhydrous($MgSO_4$) distilled water up to 1000 ml
Ampicillin	Ampicillin sodium salt was dissolved in double distilled water to a final concentration of 100 mg/ml.
IPTG (Thermo Scientific)	IPTG was dissolved in double distilled water to the final concentration of 0.1 M.
X-Gal (Thermo Scientific)	Ready to use, concentration of 20 mg/ml.

2.1.6 Solutions for DNA work

The following solutions and chemicals were used in this study.

***Aphanomyces* DNA**	0.2 M Tris hydrochloride pH 8.5
extraction buffer	0.25 M sodium chloride (NaCl)
	25 mM ethylenediaminetetraacetic acid disodium salt (EDTA)
	0.5 % sodium dodecyl sulfate (SDS)
	distilled water up to 1000 ml

The components of the DNA extraction buffer play an important role in the disruption of cell membranes while protecting the DNA from degradation. SDS, an anionic surfactant, disrupts membranes and lyses cells by binding lipids and denaturing membrane proteins; EDTA is a chelating agent that protects the DNA while disrupting the membranes by chelating divalent cations such as calcium and magnesium, which are needed to preserve the integrity of membranes and used by DNAases to degrade the DNA released from nuclei and mitochondria; Tris hydrochloride protects the DNA by acting as a buffer and stabilising the pH; sodium chloride is a salt which removes proteins bound to the DNA and neutralises the DNA negative charges, helping the DNA to precipitate in the aqueous phase (Sambrook and Russell, 2012).

Tris-equilibrated phenol	premixed, ready to use.
Chloroform:isoamyl	premixed, ready to use.
alcohol 24:1	
RNase A (Thermo Scientific)	premixed, concentration of 10 mg/ml.
Isopropanol	premixed, ready to use.

Tris-equilibrated phenol and chloroform:isoamyl alcohol are used in DNA extractions to remove proteins and lipids (organic phase) and to keep the DNA in the aqueous phase. The DNA is then purified from residual RNA by the addition of RNase A, precipitated by addition of isopropanol and washed with 70% ethanol. Alcohols, cause the DNA molecules to separate from the aqueous phase and precipitate at the bottom of the collection tube (Sambrook and Russell, 2012).

TAE electrophoresis buffer	18.6 g EDTA
(50x stock solution)	57.1 ml glacial acetic acid
	242 g Tris-base
	900 ml distilled water
	pH 8

DNA Ladders

The following ladders were used in this study: Bioline HyperLadder I 1 kb and HyperLadder II 50 bp (Figure 2.1 A and B), New England Biolabs Quick-Load Purple 1 kb DNA Ladder (Figure 2.1 C) and Promega 100bp DNA Ladder (Figure 2.1 D).

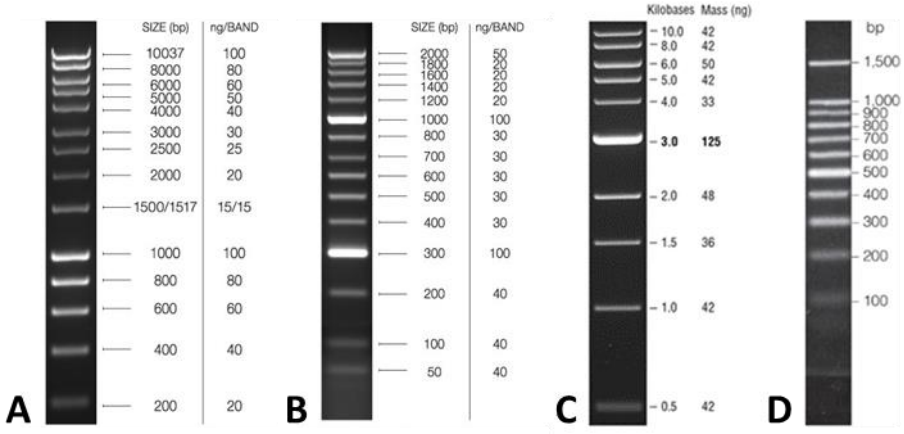


Figure 2.1 DNA ladders used in this study. A: Bioline HyperLadder I 1 kb; B: HyperLadder II 50 bp; C: New England Biolabs Quick-Load Purple 1 kb DNA Ladder; D: Promega 100bp DNA Ladder. A and B, ladders visualised by ethidium bromide staining on a 1 % TAE agarose gel when 5 µl marker are loaded (images from Bioline website: http://www.bioline.com/uk/); C, ladder visualised by ethidium bromide staining on a 0.8 % TAE agarose gel when 5 µl marker are loaded (image from New England Biolabs website: https://www.neb.com); D, ladder visualised by ethidium bromide staining on a 2 % TAE agarose gel when 5 µl marker are loaded (image from Promega website: https://www.promega.co.uk/).

2.1.7 Oomycetes used for specificity test

Genomic DNA extractions of the oomycetes listed in Table 2.3 provided by Cefas OCC were used to test the specificity of the SNV-restriction digestion assay and the genotype-specific genomic regions assays developed in this study.

Table 2.3 List of oomycetes isolates used to test the specificity of the genotyping assays developed in this study. DNA extracted and provided by Cefas OCC. Isolates numbers are indicated when known. - Unknown information.

Oomycetes isolates	Isolate ID
Achlya racemosa	-
Aphanomyces laevis	7093
Leptolegnia caudata	-
Phoma-like	-
Pythium monospermum	-
Pythium flevoense	-
Saprolegnia furcata	1-12
Saprolegnia parasitica	RP5

2.1.8 Cefas historical samples and Italian outbreak samples used to validate the genotyping methods

To test the sensitivity of the *A. astaci* genotyping methods developed in this study (sections 3.3 and 3.4), the sets of primers designed were tested on DNA extractions from infected European and North American crayfish tissues. Samples from European crayfish species were largely from white clawed crayfish (*A. pallipes*) that had been sampled from sites of mortality events of native crayfish. North American crayfish samples had been from various locations in water courses in England and Wales. Tissue samples were DNA extracted following the methods described in Tuffs and Oidtmann (2011). The outbreaks samples had tested positive for *A. astaci* at Cefas by PCR using primers 42 and 640 and sequencing of PCR products (OIE, 2016a). The carrier samples were tested positive for *A. astaci* at Cefas by qPCR, following Vrålstad *et al.* (2009) (James *et al.,* 2017). For a complete list of samples tested, see Table 5.1 and Table 5.2. Determination of agent levels follows Vrålstad *et al.* (2009).

To assess if the genotyping methods developed in the present study would produce identical results to microsatellite and RAPD-PCR genotyping techniques, samples from a well characterised Italian outbreak were tested

(Pretto *et al.*, 2014) (results in section 3.5). The samples included DNA extracts from white clawed crayfish (*A. pallipes*) and from *A. astaci* pure cultures isolated during the outbreak. These extractions had tested positive for *A. astaci* by PCR using primers 42 and 640 and sequencing of PCR products (OIE, 2016a) and qPCR (Vrålstad *et al.*, 2009). The isolate was identified by Pretto *et al.* (2014) to belong to genotype A by RAPD-PCR and microsatellites. Samples tested are listed in Table 5.3.

2.2 Microbiological methods

2.2.1 *Aphanomyces* routine culturing method

Pure cultures of *Aphanomyces* isolates (Table 2.1) were kept in axenic conditions and subcultured in GP broth and agar plates at the University of Exeter Biocatalysis Centre, following Lilley *et al.* (1998). All broth cultures were kept in 500 ml borosilicate conical flasks in a 15 °C incubator. *A. invadans* and *A. invadans*-like GP cultures were kept in the 15 °C incubator, while *A. astaci* and A. *frigidophilus* GP cultures were kept at 4 °C. *A. astaci* plates were subcultured monthly, while *A. frigidophilus, A. invadans and A. invadans*-like were subcultured fortnightly or when the mycelia growing edge reached the sides of the plate. All broth cultures were subcultured once a month. Subculturing was carried out in aseptic conditions in a Faster BH-EN and BHG laminar flow cabinet. New GP broth and agar plates were firstly removed from storage and brought to room temperature. In the cabinet, the top of empty autoclaved 500 ml flasks was flamed on a Bunsen burner and filled with 200 ml of new autoclaved GP broth. When subculturing from flasks, the flasks were gently tilted, the mycelium removed with sterile (flamed) tweezers and placed in an empty sterile plate. A portion of about 1 cm of mycelium was cut with sterile (flamed) forceps from the growing edge and placed in the previously prepared new conical flask and onto new GP agar plates. The plates and flask were then sealed with Parafilm tape on top of autoclaved aluminium foil and stored at the isolates corresponding temperatures. When subculturing from GP agar plates, a portion of about 1 cm^2 of growing mycelium was cut directly from the plate and placed in a new conical flask and new GP agar plates. Plates and flasks were

then sealed and stored at the isolates corresponding temperatures. Cultures were checked daily for contamination.

2.2.2 Preparation of LB-ampicillin broth and plates

LB-ampicillin broth was made by adding ampicillin stock solution to autoclaved LB broth to a final concentration of 100 µg/ml. For LB-ampicillin plates, ampicillin stock solution was added to autoclaved LB agar once the media was cooled down to ~50 °C to a final concentration of 100 µg/ml, the agar spread in sterile plates and left to solidify. The broth and plates were stored at 4 °C until required.

2.2.3 Transformation of *E. coli* competent cells with plasmid DNA

E. coli DH5α competent cells (Invitrogen) were slowly defrosted on ice. For each transformation, 50 µl of competent cells were transformed with 1 to 5 µl of plasmid DNA (prepared following section 2.3.6). The cells were gently flicked and left on ice for 30 minutes, heat-shocked at 42 °C for 45 seconds and incubated for two minutes on ice. 950 µl of SOB were added to the cells and the suspension was incubated at 37 °C in a shaking incubator for one hour. The suspension was then centrifuged in a bench-top centrifuge, 900 µl of supernatant discarded, the cell pellet resuspended in the remaining supernatant and spread onto LB-ampicillin agar plates. The plates were incubated overnight at 37 °C.

2.2.4 Storage of bacterial clones

Glycerol stocks of important clones were prepared and stored at -80 °C. To do this, 200 µl of glycerol was distributed in 1.5 ml microcentrifuge tubes and autoclaved at 121 °C for 15 minutes. 800 µl overnight culture of the desired clone were then added to the autoclaved glycerol and the sample stored at -80 °C.

2.3 Molecular biology methods

2.3.1 Preparation of agarose gels and gel electrophoresis

All agarose gels were made with 1x TAE electrophoresis buffer at a final concentration of 1 %, unless otherwise stated. Genomic DNA and PCR products were visualised using ethidium bromide or Midori Green Advance 20,000x. The gels were placed into the electrophoresis tank and run in a Bio-Rad wide mini-sub cell GT gel apparatus with a Bio-Rad PowerPac basic power supply. The gels were visualised in an UV light box.

2.3.2 Isolation and quantification of *Aphanomyces* genomic DNA

High molecular weight genomic DNA (gDNA) was extracted from pure *Aphanomyces* cultures (Table 2.1) following an adaptation of the protocol described in Jiang *et al.* (2013). Briefly, 7 to 14 days old mycelium was harvested under aseptic conditions from conical flasks, placed on filter paper to drain the residual GP broth and dried for 5 minutes under a laminar flow cabinet. One gram of mycelium was weighted and grinded with aid of autoclaved glass beads, pestle, and mortar. The homogenised material obtained was transferred in a sterile tube and 2 ml of *Aphanomyces* DNA extraction buffer (section 2.1.6), 2.5 ml of tris-equilibrated phenol and 1 ml of chloroform:isoamyl alcohol 24:1 were added to the homogenised material, mixed evenly and incubated at room temperature on a rocker for one hour and centrifuged at 6,000 g for 30 minutes. The aqueous phase was extracted with equal volume of chloroform:isoamyl alcohol 24:1 and centrifuged at 10,000 g for 15 minutes. 20 µl of RNase A (10 mg/ml) were added to the aqueous phase and incubated at 37 °C for 10 minutes. 0.6 volumes of isopropanol were added, mixed gently by inversion, followed by precipitation of DNA on ice for 30 minutes. The DNA was collected by centrifugation at 10,000 g for 20 minutes, the pellet washed with 70 % ethanol twice and air dried overnight at 4 °C. The DNA was then resuspended in 100 µl of DNase/RNase-free water, incubated for 15 minutes at 37 °C and checked for quality and RNA contamination by gel electrophoresis using 0.8 % agarose gel with ethidium bromide or Midori Green Advance stains. 5 µl of gDNA were mixed with 5x DNA loading buffer (Bioline)

and loaded onto the gel. Quantification of gDNA was carried out using a Qubit® 2.0 Fluorometer (Invitrogen).

2.3.3 Whole-genome sequencing of *Aphanomyces* isolates

Genomic DNA from *Aphanomyces* isolates (Table 2.4) was submitted for WGS to the Exeter Sequencing Service (University of Exeter) and sequenced using the ultra-high-throughput Illumina HiSeq 2000 System to generate paired 100 bp or 150 bp reads (Table 2.4).

Table 2.4 List of *Aphanomyces* isolates used for WGS. *Genotypes are here defined as per the Cefas OCC record system.

Species	Isolate-(Genotype)*	Read length (bp)
A. astaci	Da (A)	100
	SV (A)	100
	457 (B)	150
	197901 (B)	150
	D2 (B)	100
	SA (B)	100
	Si (B)	100
	YX (B)	100
	KV (C)	100
	Pc (D)	100
	KB13 (E)	200
A. frigidophilus	AP5	150
	RP2	150
A. invadans	NJM 0002	150
A. invadans-like	NJM 9510	150

2.3.4 PCR protocols for amplifications of targeted regions

PCR reactions were prepared under an UV2 PCR workstation (UVP, Analytik Jena) and carried out in an Eppendorf Mastercycler nexus GSX1 or a Bio-Rad T100 thermal cycler. Gradient PCR was performed to obtain the correct annealing temperatures of primers designed in this study. Negative controls without template were run for each experiment to check contamination of reagents. For the complete list of primers, see Table 2.2. PCR products were checked on agarose gels as described in section 2.3.1.

2.3.4.1 Amplifications of the internal transcribed spacer of the ribosomal DNA (ITS)

To check the identity of isolates used in this study (Table 2.1) and to test the variability of the ITS region, the primers ITS5 and ITS4 were used to amplify the ITS region. Primers ITS1 and ITS4 were instead used to amplify the same ITS region for the isolates listed in Table 2.3. PCR cycling conditions are described in Table 2.5.

ITS5-4 PCR reaction 10 µl 2x GoTaq® Hot Start Green Master Mix

1 µl 10 µM primer ITS5

1 µl 10 µM primer ITS4

1 µl gDNA template

7 µl nuclease-free water

ITS1-4 PCR reaction 10 µl 2x GoTaq® Hot Start Green Master Mix

1 µl 10 µM primer ITS1

1 µl 10 µM primer ITS4

1 µl gDNA template

7 µl nuclease-free water

Table 2.5 ITS PCR cycling conditions. Amplification with primers ITS5 and ITS4 following Diéguez-Uribeondo *et al.* (2009) and primers ITS1 and ITS4 modified protocol from Makkonen *et al.* (2011).

ITS5-4 PCR cycling conditions			
Step	**Temperature**	**Time**	**Number of cycles**
Initial denaturation	94 °C	5 min	1
Denaturation	94 °C	30 sec	5
Annealing	55 °C	30 sec	
Extension	72 °C	1 min	
Denaturation	94 °C	30 sec	33
Annealing	48 °C	30 sec	
Extension	72 °C	1 min	
Final extension	72 °C	10 min	1
ITS1-4 PCR cycling conditions			
Step	**Temperature**	**Time**	**Number of cycles**
Initial denaturation	95 °C	3 min	1
Denaturation	94 °C	1 min	35
Annealing	52 °C	45 sec	
Extension	72 °C	30 sec	
Final extension	72 °C	5 min	1

2.3.4.2 Genotyping *A. astaci* isolates by RAPD-PCR

Genomic DNA from *A. astaci* isolates was subjected to RAPD-PCR with B01 primer (Table 2.2). PCR cycling conditions are described in Table 2.6.

RAPD-PCR reaction 10 µl 2x GoTaq® Hot Start Green Master Mix

2 µl 10 µM primer B01

1 µl gDNA template

7 µl nuclease-free water

Table 2.6 RAPD-PCR cycling conditions. Protocol followed Huang *et al.* (1994).

RAPD-PCR cycling conditions			
Step	**Temperature**	**Time**	**Number of cycles**
Initial denaturation	95 °C	10 min	1
Denaturation	94 °C	1 min	45
Annealing	36 °C	1 min	
Extension	72 °C	2 min	
Final extension	72 °C	7 min	1

Amplified DNA was checked by gel electrophoresis on a 1.4 % TAE agarose gel stained with 2 µl of Midori Green Advance 20,000x. RAPD-PCR profiles were visually compared to the reference isolates available in the literature (Huang *et al.*, 1994; Kozubíková, *et al.*, 2011).

2.3.4.3 Amplifications of the cytochrome oxidase subunit I (COI)

To detect polymorphisms able to distinguish different *A. astaci* genotypes, COI was amplified with primers Oom-COI-Lev-up and Fm85mod. PCR cycling conditions are described in Table 2.7.

COI PCR reaction 10 µl 2x GoTaq® Hot Start Green Master Mix

1 µl 10 µM primer Oom-COI-Lev-up

1 µl 10 µM primer Fm85mod

1 µl gDNA template

7 µl nuclease-free water

Table 2.7 COI PCR cycling conditions. Protocol followed Robideau *et al.* (2011).

COI PCR cycling conditions			
Step	**Temperature**	**Time**	**Number of cycles**
Initial denaturation	95 °C	2 min	1
Denaturation	95 °C	1 min	35
Annealing	55 °C	1 min	
Extension	72 °C	1 min	
Final extension	72 °C	10 min	1

2.3.4.4 Amplifications of the large subunit ribosomal DNA (LSU)

The universal eukaryotic primers UNup18S42 and UN-lo28S1220 were used to amplify the LSU region. PCR cycling conditions are described in Table 2.8.

LSU PCR reaction

10 µl 2x KOD Hot Start DNA polymerase (Merck Millipore)

1 µl 10 µM primer UNup18S42

1 µl 10 µM primer UN-lo28S1220

1 µl gDNA template

7 µl nuclease-free water

Table 2.8 LSU PCR cycling conditions. Protocol followed Robideau *et al.* (2011).

LSU PCR cycling conditions			
Step	**Temperature**	**Time**	**Number of cycles**
Initial denaturation	95 °C	3 min	1
Denaturation	95 °C	30 sec	40
Annealing	68 °C	45 sec	
Extension	72 °C	2 min	
Final extension	72 °C	8 min	1

2.3.4.5 Amplifications of genotype-specific *A. astaci* genomic regions

To amplify the genotype-specific regions of *A. astaci* DNA detected with the bioinformatics approach described in section 2.4.10, the following primers were used:

- A_unique_F and A_unique_R: amplification of genotype-specific DNA region present in genotype A isolates (primer pair referred as A_unique);
- B_unique_F and B_unique_R: amplification of genotype-specific DNA region present in genotype B isolates (primer pair referred as B_unique);
- C_unique_F and C_unique_R: amplification of genotype-specific DNA region present in genotype C isolates (primer pair referred as C_unique);

- D_unique_F and D_unique_R: amplification of genotype-specific DNA region present in genotype D isolates (primer pair referred as D_unique);
- E_unique_F and E_unique_R: amplification of genotype-specific DNA region present in genotype E isolates (primer pair referred as E_unique).

PCR cycling conditions are described in Table 2.9.

"Unique regions" 10 µl 2x GoTaq® Hot Start Green Master Mix

PCR reaction 1 µl 10 µM forward primer

1 µl 10 µM reverse primer

1 µl gDNA template

7 µl nuclease-free water

Table 2.9 Genotype-specific genomic regions PCR cycling conditions. *Annealing temperatures: A_unique and C_unique: 61 °C; B_unique and E_unique: 63 °C; D_unique: 50 °C.

"Unique regions" PCR cycling conditions			
Step	**Temperature**	**Time**	**Number of cycles**
Initial denaturation	95 °C	2 min	1
Denaturation	95 °C	1 min	35
Annealing*	50/61/63 °C	45 sec	
Extension	72 °C	45 sec	
Final extension	72 °C	5 min	1

2.3.4.6 Amplifications of *A. astaci* genotype-specific SNVs in nuclear region

To amplify the regions containing *A. astaci* genotype-specific SNVs in the nuclear DNA detected with the bioinformatics approach and digestible by HhaI restriction enzyme (sections 2.4.8 and 2.4.9), the following primers were used:

- A_HhaI_F and A_HhaI_R: amplification of nuclear region with SVN present in genotype A isolates (primer pair referred as A_HhaI);
- B_HhaI_F and B_HhaI_R: amplification of nuclear region with SVN present in genotype B isolates (primer pair referred as B_HhaI);
- C_HhaI_F and C_HhaI_R: amplification of nuclear region with SVN present in genotype C isolates (primer pair referred as D_HhaI);
- D_HhaI_F and D_HhaI_R: amplification of nuclear region with SVN present in genotype D isolates primer pair referred as C_HhaI);

- E_HhaI_F and E_HhaI_R: amplification of nuclear region with SVN present in genotype E isolates (primer pair referred as E_HhaI).

PCR cycling conditions are described in Table 2.10.

"HhaI" PCR reaction 10 µl 2x GoTaq® Hot Start Green Master Mix

1 µl 10 µM forward primer

1 µl 10 µM reverse primer

1 µl gDNA template

7 µl nuclease-free water

Table 2.10 "HhaI" PCR cycling conditions. *Annealing temperatures: A_HhaI, B_HhaI and C_HhaI: 60 °C; D_HhaI: 58 °C; E_HhaI: 59 °C.

"HhaI" PCR cycling conditions			
Step	**Temperature**	**Time**	**Number of cycles**
Initial denaturation	95 °C	2 min	1
Denaturation	95 °C	1 min	35
Annealing*	58/59/60 °C	45 sec	
Extension	72 °C	45 sec	
Final extension	72 °C	5 min	1

2.3.4.7 Amplification of *A. astaci* genotype-specific SNVs in nuclear region and genotype-specific *A. astaci* regions with touchdown PCR

To improve the amplification of *A. astaci* targeted nuclear DNA region from Cefas outbreak and carriers' samples (2.1.8), touchdown PCR was applied on a subset of samples. The following primer pairs were used:

- A_HhaI, B_HhaI, C_HhaI, D_HhaI, E_HhaI;
- A_unique, B_unique, C_unique, D_unique, E_unique.

PCR cycling conditions are described in Table 2.11.

Touchdown 10 µl 2x GoTaq® Hot Start Green Master Mix

PCR reaction 1 µl 10 µM forward primer

1 µl 10 µM reverse primer

1 µl gDNA template

7 µl nuclease-free water

Table 2.11 Touchdown PCR cycling conditions. *Annealing temperatures were set three degrees higher than the optimised PCR cycling conditions (Table 2.9 and Table 2.10) and decreased by one degree per cycle. **Annealing temperatures in the last 15 cycles were set 15 degrees lower than the optimised PCR cycling conditions (Table 2.9 and Table 2.10).

Touchdown PCR cycling conditions			
Step	**Temperature**	**Time**	**Number of cycles**
Initial denaturation	95 °C	2 min	1
Denaturation	95 °C	1 min	30
Annealing*	*	1 min	
Extension	72 °C	45 sec	
Denaturation	95 °C	1 min	15
Annealing**	**	1 min	
Extension	72 °C	45 sec	
Final extension	72 °C	5 min	1

2.3.4.8 Amplifications of *A. astaci* genotype-discriminatory SNVs in mitochondrial region

The following primers were used to amplify the regions containing SNVs that discriminate between the previously defined RAPD-PCR-based genotypes in the mtDNA:

- A_MITO_F and A_MITO_R: amplification of mitochondrial region with SVN present in genotype A isolates (primer pair referred as A_MITO);
- B_MITO_F and B_MITO_R: amplification of mitochondrial region with SVN present in genotype B isolates (primer pair referred as B_MITO);
- C_MITO_F and C_MITO_R: amplification of mitochondrial region with SVN present in genotype C isolates (primer pair referred as C_MITO);
- D_MITO_F and D_MITO_R: amplification of mitochondrial region with SVN present in genotype D isolates (primer pair referred as D_MITO);
- E_MITO_F and E_MITO_R: amplification of mitochondrial region with SVN present in genotype E isolates (primer pair referred as E_MITO).

PCR cycling conditions are described in Table 2.12.

Mitochondrial target 10 µl 2x GoTaq® Hot Start Green Master Mix

PCR reaction 1 µl 10 µM forward primer

1 µl 10 µM reverse primer

1 µl gDNA template

7 µl nuclease-free water

Table 2.12 Mitochondrial SNV target PCR cycling conditions. *Annealing temperatures: A_MITO: 58 °C; B_MITO, C_MITO and D_MITO: 59 °C; E_MITO: 55 °C.

Mitochondrial regions PCR cycling conditions			
Step	**Temperature**	**Time**	**Number of cycles**
Initial denaturation	95 °C	2 min	1
Denaturation	95 °C	1 min	35
Annealing*	55/58/59 °C	45 sec	
Extension	72 °C	45 sec	
Final extension	72 °C	5 min	1

2.3.4.9 Semi-nested PCR protocol for detection of genotype B

A semi-nested PCR assay was developed to exploit the SNVs present in the region amplified by B_MITO primer pair (sections 2.3.4.8). For this genotype, it was not possible to find a restriction enzyme able to digest the PCR product at the level of the SNV. The following primers were used:

- B_MITO_N1S and B_MITO_R: amplification of mitochondrial region with SVN present in genotype B isolates;
- B_MITO_N2S and B_MITO_R: amplification of mitochondrial region with SVN absent in genotype B isolates, used as control primers.

The forward primer N1S was designed before the SNV of interest, with last base matching the SNV. The forward primer N2S was designed in the same position as N1S, but with different last (*i.e.* 3') base (Figure 2.2).

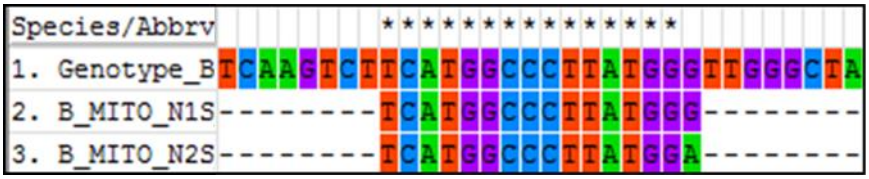


Figure 2.2 Semi-nested forward primers aligned to genotype B fragment. * matching nucleobases. N1S and N2S differ in the 3' base.

Amplifying the PCR product obtained from the first round PCR (following protocol section 2.3.4.8, primer pair B_MITO) with the semi-nested primers, the presence or absence of a second band of smaller size in comparison to the original PCR product on the gel indicates that the primers matched the variant in the template (Chen and Sullivan, 2003). By using one nested primer at a time

in the PCR reaction, the PCR product can be ascribed to genotype B or not. PCR cycling conditions are described in Table 2.13.

Semi-nested	10 µl 2x GoTaq® Hot Start Green Master Mix
PCR reaction	1 µl 10 µM forward primer (B_MITO_N1S or B_MITO_N2S)
	1 µl 10 µM B_MITO_R
	1 µl template (PCR product from B_MITO PCR)
	7 µl nuclease-free water

Table 2.13 Semi-nested PCR cycling conditions. *Annealing temperatures: B_MITO_N1S: 65 °C; B_MITO_N2S: 62.8 °C.

Semi-nested PCR cycling conditions			
Step	**Temperature**	**Time**	**Number of cycles**
Initial denaturation	95 °C	2 min	1
Denaturation	95 °C	1 min	20
Annealing*	62.8/65 °C	45 sec	
Extension	72 °C	30 sec	
Final extension	72 °C	5 min	1

2.3.5 Purification of PCR reactions

PCR reactions showing a single band on agarose gels were purified with GeneJET PCR Purification Kit (Thermo Scientific). For LSU fragments, as more than one band was present on the gel, the bands of expected size were excised with a scalpel and extracted with QIAquick Gel Extraction Kit (Qiagen). Purifications followed the manufacturers' instructions. PCR concentrations were estimated using NanoDrop 2000 spectrophotometer (Thermo Scientific) and samples were stored at -20 °C.

2.3.6 Cloning into pGEM-T Easy (Promega) and pJET1.2 (Thermo Scientific) vectors

COI, ITS, and LSU purified PCR products from isolates listed in Table 2.1 and RAPD-PCR purified PCR products from *A. astaci* isolates 457 and D2 were ligated into pGEM-T Easy (Promega) or pJET1.2 (Thermo Scientific) vectors. All the ligations were performed at a 3:1 ratio insert to vector, following

manufacturers' protocols. Vectors were transformed into *E. coli* DH5α competent cells following section 2.2.3.

2.3.7 Colony screening and plasmid extraction

Following the cloning of PCR products (section 2.3.6) and the transformation of plasmids (section 2.2.3), overnight colonies were screened by PCR following manufacturer's protocol. PCR products were checked by gel electrophoresis, positive clones were incubated overnight at 37 °C in LB-ampicillin broth at 150 rpm for plasmid extraction. Plasmid DNA was extracted and purified using QIAprep Spin Miniprep Kit (Qiagen), checked by PCR and gel electrophoresis as described above, quantified with a NanoDrop 2000 spectrophotometer (Thermo Scientific) and stored at -20 °C.

2.3.8 Sequencing of purified PCR products and cloned PCR products

PCR products and cloned PCR products of interest were sent to Eurofins MWG Operon commercial sequencing facility to sequence with the Sanger sequencing method. PCR products of *A. astaci* regions containing genotype-specific SNVs and genotype-specific regions were sequenced with their respective amplification primers. ITS PCR products of the oomycetes listed in Table 2.3 were sequenced with primer ITS4. For plasmids containing COI, ITS and LSU fragments, between 3 and 6 clones from each isolate (Table 2.1), were sequenced with the primers associated with the plasmid and provided with the cloning kit used. RAPD-PCR cloned fragments were sequenced with the primers associated with the plasmid and provided with the cloning kit used. For a complete list of primers see Table 2.2.

2.3.9 Enzymatic digestion of PCR products for genotyping *A. astaci*

One of the protocols developed in this study to detect specific *A. astaci* genotypes involved the detection of genotype-specific SNVs contained in a restriction digestion site. Thus, by amplifying a specific region of the DNA by PCR and digesting the PCR product, the isolate's genotype could be determined by the pattern of bands on the agarose gel. After amplification of the

region containing the genotype-specific SNV, an enzymatic digestion was applied with the following restriction enzymes:

- Region amplified by A_MITO primer pair: enzyme MspI;
- Region amplified by C_MITO primer pair: enzyme MseI;
- Region amplified by D_MITO primer pair: enzyme AciI;
- Region amplified by E_MITO primer pair: enzyme Taq$^{\alpha}$I;
- Region amplified by A_HhaI, B_HhaI, C_HhaI, D_HhaI and E_HhaI primer pairs: enzyme HhaI.

Prior to digestion, PCR reactions were checked for presence of product and quantified on 1% agarose gel, stained with Midori Green Advance stain (section 2.3.1). All restriction enzymes used in this study were obtained from New England Biolabs and digestions followed the manufacturer protocol: 300 ng of non-purified PCR product were mixed with 2 µl of 10x CutSmart Buffer (New England Biolabs), 0.3 µl of enzyme and nuclease-free water up to 20 µl. The mixture was incubated following manufacturer instructions at 37 °C for 1 (AciI, MseI and MspI) or 2 hours (HhaI) or at 65 °C for 1 hour (Taq$^{\alpha}$I). Digested products were then checked by gel electrophoresis using 2 % agarose gel stained with Midori Green Advance stain.

2.4 Bioinformatics methods

2.4.1 Cloned sequences analysis and alignments

Sanger sequencing reads (COI, ITS, LSU and RAPD-PCR cloned sequences) and their associated chromatograms were reviewed using BioEdit version 7.0.8 (Hall, 1999). For each isolate, a single consensus sequence was generated combining the respective cloned sequences using BioEdit, with threshold frequency value set at 95 % and gaps treated like residues. Nomenclature for ambiguous nucleotides follows the IUPAC nomenclature code (Cornish-Bowden, 1985).

2.4.2 Phylogenetic analysis of cloned sequences

For COI, ITS, and LSU phylogenetic analyses were undertaken. The National Centre for Biotechnology Information (NCBI) Basic Local Alignment Search Tool (BLAST) search was used to obtain previously available homologous sequences from the NCBI Nucleotide database. In addition to the 20 isolates held in the laboratory (Table 2.1), 36 ITS sequences, 15 LSU sequences and 6 COI sequences, for a total of 12 named *Aphanomyces* species and two strains not assigned to a named species were retrieved from the NCBI Nucleotide database. *Saprolegnia parasitica* and *A. invadans* were used as outgroups. Complete lists of sequences retrieved with species, isolate ID, and GenBank accession numbers are provided in Table 5.4, for COI, Table 5.5 for ITS, Table 5.6 for *A. astaci* ITS phylogenetic analysis, and Table 5.7 for LSU. *A. astaci* genotypes of retrieved sequences and references are indicated where necessary. MUSCLE (Edgar, 2004) was used to construct multiple sequence alignments for each amplicon, including homologous sequences identified from the BLAST searches. The alignments were viewed using MEGA6 software (Tamura *et al.*, 2013). Consensus sequences were then aligned using MUSCLE and the alignments between different isolates were visualised using MEGA6 software. Sequence differences between isolates were noted. Phylogenetic reconstruction of the sequences was performed with MEGA6 using a maximum likelihood statistical analysis. Robustness of phylogeny was tested using the bootstrap method, with 500 replications. Gaps were treated with the Partial-Deletion option and Site Coverage Cutoff set at 95 %. This option removes sites of deletion before the phylogenetic analysis begins if the site has an ambiguity percentage value higher that the Site Coverage Cutoff parameter.

2.4.3 Quality control of genome-wide sequencing data

For each sequenced isolate (Table 2.4), Illumina HiSeq 2000 pair-end reads were quality checked with FastQC (Andrews, 2010) version 0.11.3. FastQC is a bioinformatics tool to check the quality of raw sequence data from high throughput sequencing systems. It provides an intuitive set of analyses that gives a good indication of the quality of the data.

Trim Galore! (Krueger, 2012) version 0.3.7 was run on the paired reads to remove reads of low-quality, adapter sequences from the 3' end of the reads and filter trimmed reads based on their sequence length. It uses the adapter trimming tool Cutadapt (Martin, 2011) version 1.7.1 and FastQC to quality trim reads. This step is important to remove poor quality and short reads that can affect the genome assembly process. Trim Galore! was run with default settings and option '--paired' on. This option operates a paired-end validation on the trimmed reads, removing both read pairs if at least one of the two paired sequences is shorter than the other one. For *A. astaci* KB13 pair-end reads, the FASTX-Toolkit (http://hannonlab.cshl.edu/fastx_toolkit/) was used on one of the libraries to remove reads shorter than 200 bp before using Trim Galore! as the quality of these reads was still low with only Trim Galore!.

2.4.4 *De novo* assembly of *Aphanomyces* genomes

Numerous form of software are available for *de novo* assembly of Illumina sequence data, amongst the most widely used being Velvet (Zerbino and Birney, 2008), SOAPdenovo (Luo *et al.*, 2012) and SPAdes (Bankevich *et al.*, 2012). These three form of software initially build a de Bruijn graph based on a given *k*-mer length by hashing paired-end reads. Once the graph is set, the reads paths are traced back along the graph and, following the software specific computational analysis, contigs and scaffolds are constructed. Short descriptions of software used in this study are detailed below.

The performance of all three form of software cited above was tested on WGS data from a subset of sequenced isolates in order to choose the optimal method. Paired-end reads from *A. astaci* 457 and 197901, *A. frigidophilus* AP5 and RP2, *A. invadans* NJM0002 and *A. invadans*-like NJM9510 were run on three different assemblers: SOAPdenovo, Velvet and SPAdes. The obtained assemblies were then assessed for completeness with the Benchmarking Universal Single-Copy Orthologs (BUSCO) software (Simão *et al.*, 2015) (2.4.5).

- SOAPdenovo2 (version 2.04-r240):

SOAPdenovo2 is a *de novo* genome assembler for short reads. It runs six sections that manage read error correction, construction of the de Bruijn graph, assembly of contigs and scaffolds, mapping of paired end reads and gap closure (through GapCloser). SOAPdenovo2 was run on the paired reads of *A. frigidophilus* AP5 and *A. invadans* NJM002 following the manual instructions and default parameters, with all sets of *k*-mer lengths and with the option to resolve repeats on. The values for maximal read length and the average insert size were set to 150 and 500 respectively.

- Velvet (version 1.2.10):

Velvet is a *de novo* genome assembler designed for short reads which is based on the de Bruijn graph, constructed by hashing the reads per a given *k*-mer length. Once the graph is set, the reads paths are traced along the graph. After constructing the de Bruijn graph, it also removes errors from the dataset by using its algorithm Tour Bus and resolves repeats, producing contigs as final output. Velveth was run on the paired reads following the manual instructions, with all sets of *k*-mer lengths between 65 and 113, with a step of two. Velveth calculates automatically some informative values, like the expected genome coverage and the coverage cut off and uses them in the genome assembly process. To see if the assemblies could be improved, Velvetg was run on the pair-end reads following the manual instructions, with all sets of *k*-mer lengths between 65 and 113, with a step of two, but with modified coverage cut off.

- SPAdes Genome Assembler (version 3.5.0):

SPAdes is the St. Petersburg *de novo* genome assembler that, after building the de Bruijn graph, uses graph-theoretical processes based on graph topology, depth of coverage and sequence lengths to return a consensus DNA sequence. In particular, SPAdes infers distances between *k*-mers by analysing distance histograms and paths in the de Bruijn graph, then uses the paired de Bruijn graph approach to combine paired reads information with the paired assembly graph. In the last step, SPAdes builds DNA contigs, remaps the paired reads to the contigs and returns a consensus DNA sequence. SPAdes was run on the paired reads following the manual instructions, with all sets of *k*-mer lengths between 19 and 129 and with the option '--careful' on. This option reduces the number of mismatches and short indels using MismatchCorrector. On SPAdes assemblies, GapCloser (SOAPdenovo gap closer software, v1.12) was run to

close the gaps between the scaffolds generated by the assembler. Defaults parameters were used, except the average insert size ‘avg_ins’, which was calculated for each assembly by aligning the reads against the newly made assembly with the Burrows-Wheeler Aligner tool (BWA) (Li, 2013) and calculating the mean of the observed template length (TLEN). BWA is a software package that maps sequences against a reference genome. The maximal read length ‘max_rd_len’ was set based on the type of Illumina run (Table 2.4). Contigs and scaffolds less than 200 bp and 500 bp in length respectively were removed.

After assessment with BUSCO (2.4.5) on the assemblies obtained with the three assemblers, the one chosen and used for all isolates was SPAdes (version 3.5.0).

2.4.5 Quality checks on assemblies

Determining the completeness and the contiguity of an assembled genome is an important step to assess the quality of a new genome. The completeness and the contiguity of the genomes assembled in this study were assessed using BUSCO and the QUality ASsessment Tool (QUAST) (Gurevich *et al.*, 2013).

BUSCO is a software that estimates the completeness of an assembled genome by performing a census of conserved single-copy genes that ought to be found in the genomes of particular lineages of organisms. Assessing their presence in single or multi-copy and their completeness in a genome, gives an indication of how well the genome has been assembled. BUSCO software uses orthologs from OrthoDB (http://www.orthodb.org) and was run with the lineage-specific library “Eukaryota”, downloaded from http://busco.ezlab.org (Simão *et al.*, 2015).

QUAST is a tool that evaluates and compares genome assemblies producing various quality metrics giving indications on the contiguity of the assemblies (Gurevich *et al.*, 2013). In particular, the most used statistic parameters used to evaluates the contiguity of an assembly are the contigs (or the scaffolds) N_{50} number and the N_{50} length (Paszkiewicz and Studholme, 2010). N_{50} number represents the number of contigs (or scaffolds) needed to cover 50 % of the

genome. The lower is this value, the longer are the contigs (or scaffolds) that construct that assembly; the N_{50} length is the average length of contigs (or scaffolds) needed to complete 50 % of the genome, thus the higher is this value, the longer are the contigs (or scaffolds) that form the assembly. A highly contiguous assembly has a low N_{50} number and a high N_{50} length.

The assemblies were also screened for the presence of control sequencing vector sequences using *blastn*, a BLAST application that confronts nucleotides queries against a nucleotide database. The NCBI Nucleotide database was used and contigs corresponding to control sequencing vectors were removed. Control sequencing vectors are libraries derived from well characterised small genomes ligated to the adapters used in sequencing runs and used to control the performance of the run. The results of the *blastn* searches were used also to investigate the presence of unusual hits that could indicate contamination of the pure cultures with other oomycetes or bacteria.

2.4.6 *Aphanomyces astaci* mitochondrial DNA assembly

A mitochondrial genome was assembled for *A. astaci* to be used as reference genome and generate the list of *A. astaci* genotype-specific mitochondrial SNVs. A related mitochondrion genome was retrieved from NCBI (*Saprolegnia ferax,* GenBank accession number: AY534144.1) and trimmed pair-end reads of *A. astaci* D2 were aligned against it using BWA (section 2.4.7). Mapped reads were then extracted using SAMtools (Li *et al.*, 2009) *samtools view* package and assembled using SPAdes. SAMtools is a library and software package used to analyse and handle alignments in the SAM or BAM formats (Li *et al.*, 2009), BWA output files.

2.4.7 BWA alignments and SAMtools

In order to generate a list of *A. astaci* genotype-specific SNVs and genotype-specific regions, trimmed pair-end reads from each sequenced *A. astaci* isolate were aligned against a reference genome with BWA and SAMtools. The following steps were used to align the reads to the references genomes:

1. *bwa index*: creates an index of the reference genome sequences;
2. *bwa mem*: aligns the pair-end reads to the reference genome;
3. *samtools view* (with option *–bS*): prints the alignments from the input SAM file to a BAM file;
4. *samtools sort*: sorts the alignments by the leftmost coordinates;
5. *samtools index*: creates a coordinate-sorted BAM file for fast access (to be used for IGV).

The reference genome used for the genotype-specific SNVs alignments was *A. astaci* APO3 assembly, retrieved from NCBI and made available by the Broad Institute (GenBank assembly accession: GCA_000520075.1). Reference genomes used to create alignments to retrieve genotype-specific regions (section 2.4.10) were assembled in this project.

2.4.8 Detection of single nucleotide variants (SNVs)

The script developed by Dr David Studholme (Clarke *et al.*, 2015; Mazzaglia *et al.*, 2012) was used to call genotype-specific SNVs. The script categorises every genomic position as ambiguous (*e.g.* due to heterozygosity, or insufficient data) or unambiguous (*i.e.* sufficient depth and consensus of aligned sequence data) and requires a reference genome, pileup files, a “depth of coverage” and a “percentage of unambiguity” thresholds set by the operator. The reference genome used for the SNVs in nuclear regions was *A. astaci* APO3 assembly (GenBank assembly accession: GCA_000520075.1). SAMtools *mpileup* tool was used to generate a pileup file for each sorted BAM file generated following section 2.4.7. In this file format, each genomic position is listed and associated with data regarding for example reference base, read base and number of reads covering the base, information of base match/mismatch/indel. For a complete description of the format, see SAMtools manual, http://www.htslib.org/doc/samtools.html. The depth of coverage value corresponds to the minimum number of reads that cover a genomic position; the script discards the variant if the coverage for the genomic position is lower than the value. Low coverage regions can indicate misassembly or sequencing bias. The percentage of unambiguity value is the proportion of reads that

unambiguously match a variant at a genomic position. The script discards the genomic position if the proportion of reads matching the variant is below the set value. If the percentage of identity value is set high, the script discards ambiguous sites and read errors. The depth of coverage and the percentage of unambiguity thresholds were set at 10 and 95 % respectively. The script does not consider gaps and deletions and categorises each genomic position as ambiguous or unambiguous confronting the reference genome with the pileup files. It returns a list of unambiguous genomic positions, variants for each pileup input file and shared variants between pileup files, which were then used to establish genotype-specific SNVs.

2.4.9 Selection of SNVs associated with restriction sites

The script used to detect genotype-specific SNVs in restriction sites was developed by Dr David Studholme. The script requires the same reference genome used for the genotype-specific SNVs call and the output file from section. The script retrieves the SNV positions on the reference genome, recovers the nucleotide flanking sequences and constructs complementary sequences with the other nucleotide variant. It proceeds to do an *in silico* restriction digestion on the sequences and compares the list of fragments resulting from the two complementary sequences. If the lists of fragments are different, the SNVs associated with the restriction enzyme are categorised as useful to distinguish the two sequences. The script returns a list of SNVs, flanking sequences, restriction enzymes associated with the SNVs and *in silico* fragments sizes. The flanking sequences and SNVs were visualised in IGV and primers to amplify the fragments were designed.

2.4.10 Identification of sequences unique to a subset of genomes

To identify genotype-specific regions of *A. astaci* DNA, for each genotype-reference genome, trimmed pair-end reads from each sequenced *A. astaci* isolate were aligned with BWA (section 2.4.7) and BEDTools (Quinlan and Hall, 2010) *genomecov* package was used to retrieve the depth of coverage of

contigs. Contigs with zero coverage for the non-matching genotypes were selected, visualised in IGV and genotype-specific primers designed.

Reference genomes used to create alignments to retrieve genotype-specific regions were:

- *A. astaci* Da for genotype A;
- *A. astaci* D2 for genotype B;
- *A. astaci* KV for genotype C;
- *A. astaci* Pc for genotype D;
- *A. astaci* KB13 for genotype E.

2.4.11 Construction of SNV-based phylogenetic networks

To phylogenetically analyse the SNVs differences between *A. astaci* isolates and genotypes (retrieved from the WGS data with the script described in 2.4.8), the NeighborNet method embedded in SplitsTree4 (program developed by Huson and Bryant, 2006), was used. This method constructs phylogenetic split networks based on pseudosequences generated by concatenating the SNVs. Split networks are planar graphs in which the edges of the network intersect only at their ends, without crossing other edges. Every edge represents a different taxon and, similarly to the lengths of the branches in phylogenetics trees, the lengths of the edges are proportional to the weight of the associated split. Differently from phylogenetic tree-based methods (characterised by the presence of one parent node), split networks are composed by hybrid nodes (with two parent nodes), better clarifying the relationships between the sequences used in the analysis (Huson and Bryant, 2006).

2.4.12 *Aphanomyces astaci* genotypes comparisons using dnadiff

In order to assess the similarity between *A. astaci* genotypes, dnadiff (Kurtz *et al.*, 2004) was used to compare the assemblies of the 11 *A. astaci* isolates sequenced and assembled in this study and *A. astaci* APO3 sequenced and assembled by the Broad Institute (GenBank assembly accession: GCA_000520075.1). Dnadiff is a program that utilises nucmer and MUMmer packages (Kurtz *et al.*, 2004) to compare highly similar genomes (*e.g.*

comparing isolates belonging to the same species). While nucmer is used to align large DNA sequences, dnadiff processes nucmer outputs and returns a report that includes alignments statistics (*e.g.* % of aligned/unaligned sequences and % of sequence identities), which are used to understand the differences between the assemblies. Moreover, to identify the level of similarity/difference between *Aphanomyces* species, all *A. astaci* assemblies were compared to the other *Aphanomyces* assemblies generated in the present study (Table 2.4) and *A. invadans* NJM9701 sequenced and assembled by the Broad Institute (GenBank assembly accession: GCA_000520115.1).

3. Results

3.1 RAPD-PCR profiles for *Aphanomyces astaci* isolates, including *A. astaci* isolate 457, which displays an atypical RAPD-PCR profile

Aphanomyces astaci is a specialised parasite of freshwater crayfish, heavily involved in the decline of European indigenous crayfish species (Alderman *et al.*, 1984; Unestam, 1972). Applying RAPD-PCR on pure cultures has, so far, revealed five genotypes named A, B, C, D, and E (Diéguez-Uribeondo *et al.*, 1995; Huang *et al.*, 1994; Kozubíková *et al.*, 2011). Genotyping techniques for *A. astaci* include RAPD-PCR, AFLP and microsatellites with RAPD-PCR as preferred technique to use on axenic cultures (Grandjean *et al.*, 2014; Rezinciuc *et al.*, 2014). Thus, RAPD-PCR was applied to the 11 *A. astaci* isolates provided by Cefas to determine or detect intra-species differences between isolates and ascribe the isolates to genotypes.

3.1.1 Aims and objectives

The work presented below describes the RAPD-PCR profiles of the 11 *A. astaci* isolates used in subsequent parts of the thesis work and the investigation on isolate 457, which displays an atypical RAPD-PCR profile. The reader is referred to section 1.2.5 and section 1.3 for the background information on the use of genotyping techniques on *A. astaci*.

3.1.2 Materials and Methods

All methods and materials are described in detail in Chapter 2 – Materials and Methods. Briefly, gDNA was extracted from the 11 *A. astaci* isolates (Table 2.1) as described in section 2.3.2. Genomic DNA was amplified using primer B01 (Table 2.2) by RAPD-PCR (section 2.3.4.2). PCR fragments were separated by gel electrophoresis on 1.4 % agarose gel stained with Midori Green Advance stain (section 2.3.1). PCR products from isolate 457 and D2 were purified (section 2.3.5), cloned into pJET1.2 (Thermo Scientific) vector (section 2.3.6)

and transformed into *E. coli* DH5α competent cells (section 2.2.3). Colonies were screened by PCR, positive plasmids extracted with QIAprep Spin Miniprep Kit (Qiagen, Germany), sequenced (sections 2.3.7 and 2.3.8) and sequences checked with BioEdit version 7.0.8 (Hall, 1999). Using *blastn*, consensus sequences generated with BioEdit were searched against-genome assemblies of isolates 457, D2, SA, Si, YX and 197901, assembled following section 2.4.4. To investigate the features of the most interesting cloned sequences, paired-end reads from D2, SA, Si, YX, and 457 were aligned to 197901 assembly with BWA (Li, 2013) and SAMtools (Li *et al.*, 2009) (section 2.4.7) and visualised in IGV (Robinson *et al.*, 2011; Thorvaldsdóttir *et al.*, 2013).The number of *A. astaci* genotype-specific SNVs was generated by aligning trimmed pair-end reads from each sequenced *A. astaci* isolate against a reference genome with BWA and SAMtools (section 2.4.7). Reference genome was *A. astaci* APO3 assembly (GenBank assembly accession number: GCA_000520075.1). SNVs were called as per section 2.4.8. To construct a phylogenetic network based on SNVs for all *A. astaci* genotypes, the pileup files generated with SAMtools *mpileup* were analysed with SplitsTree4 (section 2.4.11).

3.1.3 RAPD-PCR profiles of *Aphanomyces astaci* isolates held in the Cefas Oomycetes Culture Collection

The patterns of the bands obtained from the RAPD-PCR for all *A. astaci* isolates (Figure 3.1) visually matched the already known and characterised genotypes (Diéguez-Uribeondo *et al.*, 1995; Huang *et al.*, 1994; Kozubíková *et al.*, 2011) and the genotypes recorded in Cefas OCC record system (Table 2.1). Isolate 457 was the only exception, whose RAPD-PCR pattern did not match any of those previously described, despite being designated as genotype B in Cefas records. *A. astaci* 457 (referred as isolate “FDL457”) was isolated from a crayfish plague outbreak on the River Arrow (Herefordshire, UK) in 1990 by D. J. Alderman (Alderman, 2003) and genotypically characterised by RAPD-PCR by (Lilley *et al.*, 1997a). Even though the gel images presented in this last study were not clear, the authors concluded that the isolate belonged to genotype B. In the current study, the RAPD-PCR profile for 457 showed a different unusual pattern in comparison to the other isolates belonging to genotype B. The bands

pattern of 457 is also different from the other previously described genotypes.

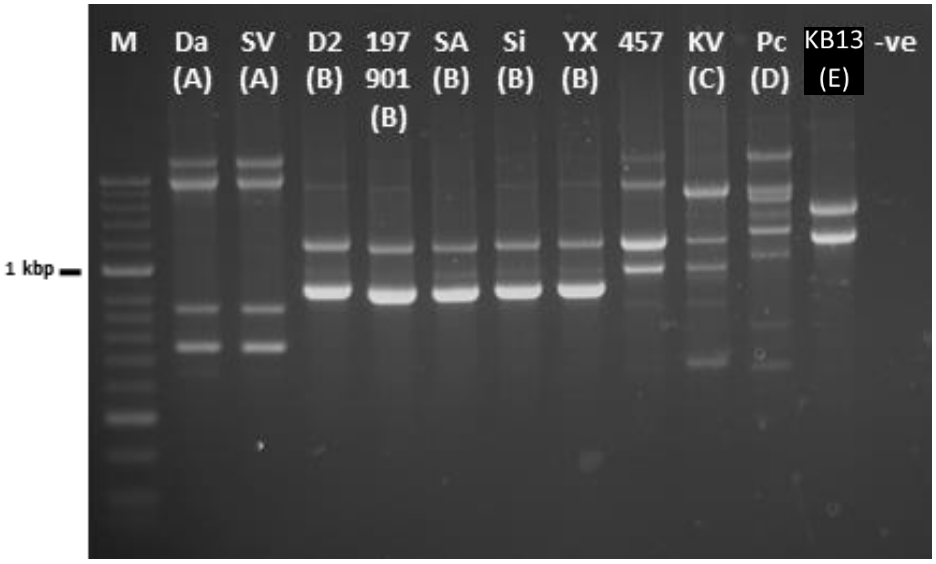


Figure 3.1 RAPD-PCR profiles of *A. astaci* isolates. Genotypes are indicated in brackets. M, Bioline HyperLadder II; -ve, negative control.

The reason for this atypical profile is investigated below.

3.1.4 Investigation of *Aphanomyces astaci* isolate 457 atypical genotype B RAPD-PCR profile by comparison of single nucleotide variants (SNVs) between genotypes

Given the unusual RAPD-PCR profile obtained for isolate 457, apparently inconsistent with its designation as a member of genotype B, a study on its phylogenetic position utilising SNVs across the whole genome (nuclear and mtDNA) was carried forward. Aligning the reads of the 11 *A. astaci* sequenced in this study to a reference assembly (*A. astaci* isolate APO3, GenBank assembly accession: GCA_000520075.1) between hundreds to several thousands of single-nucleotide polymorphisms (SNVs, *i.e.* single-nucleotide sites in the genome that are variable among the sequenced isolates) were identified and used to clarify the position of isolate 457 in the genotype classification by construction of phylogenetic networks. Table 3.1 lists the genotype-specific SNVs found in the whole genome of each genotype and the number of isolate-specific SNVs found in the genome of each specific genotype. The SNVs were concatenated to generate pseudosequences and to construct split networks, planar graphs in which every edge represents a different taxon (Figure 3.2, 3.3 and 3.4). The distances between edges represent the genetic differences between them and it is proportional to the weight associated with the split.

Table 3.1 Numbers of isolate-specific SNVs and genotype-specific SNVs. Each isolate belonging to a genotype shares the same SNVs with the isolate belonging to the same genotype. Isolate APO3 is not included in the list of isolate-specific SNVs as used as reference genome. - no isolate-specific SNVs detected.

Isolate (genotype)	Isolates SNVs	Genotypes SNVs
Da (A)	84	76'822
SV (A)	29	
D2 (B)	3	336
SA (B)	1	
Si (B)	2	
YX (B)	-	
457 (B?)	5	
197901 (B)	-	
KV (C)	1'808	1'808
Pc (D)	1'352	176'698
KB13 (E)	4'775	4'775

The first computational analysis involved all sequenced *A. astaci* isolates and the SNVs detected by aligning the paired-end reads against a reference genome (APO3, genotype D). In this phylogenetic network, isolates belonging to genotype A and D grouped in separate branches from genotypes B, C, and E with isolate 457 fitting within the genotype B group (Figure 3.2, black arrow). These results, suggest that isolates belonging to B, C, and E genotypes are closely related and that there is a higher number of genetic differences between them and genotypes A and D (Figure 3.2).

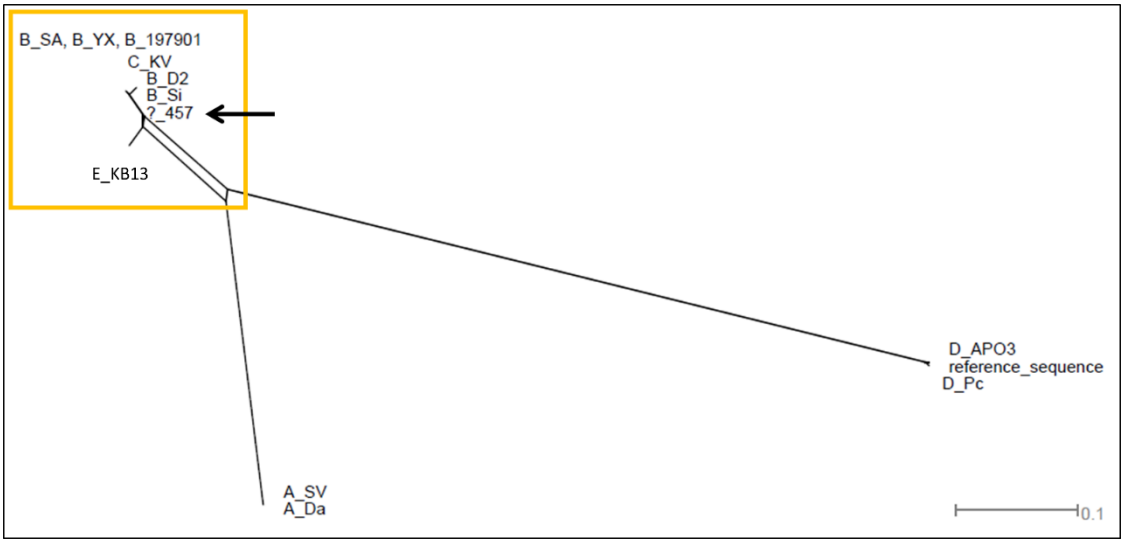


Figure 3.2 Phylogenetic network based on SNVs found in the whole genome of all *A. astaci*. The distances between *A. astaci* isolates represent the genetic differences between isolates and it is proportional to the weight associated with the split. Isolates belonging to genotype A and D grouped separately while isolates belonging to genotype B, C and E and isolate 457 grouped in the same branch (yellow box). Reference genome: *A. astaci* APO3 (genotype D).

To investigate further the relation of these closely related genotypes and to confirm the affinity of isolate 457 to genotype B, a second computational analysis involving only *A. astaci* isolates belonging to genotype B (D2, SA, Si, YX, and 197901), genotype C (KV), genotype E (KB13), and isolate 457 was carried out and a second phylogenetic network based on the SNVs detected was generated. In this analysis, *A. astaci* D2 genome assembly was used as reference genome (Figure 3.3). With this second approach the SNVs between genotypes B, C and E are sufficient to distinguish these three genotypes, confirming the genetic differences identified by RAPD-PCR.

Moreover, magnifying the genotype B branch reveals that isolate 457 is indeed more closely related to genotype B than to isolates belonging to genotype C and E (Figure 3.4). The number of SNVs reported by the computational analysis that can distinguish isolate 457 from the other genotype B isolates considered is only four, indicating that the genetic differences between 457 and genotype B are few and from this approach *A. astaci* isolate 457 belongs to genotype B.

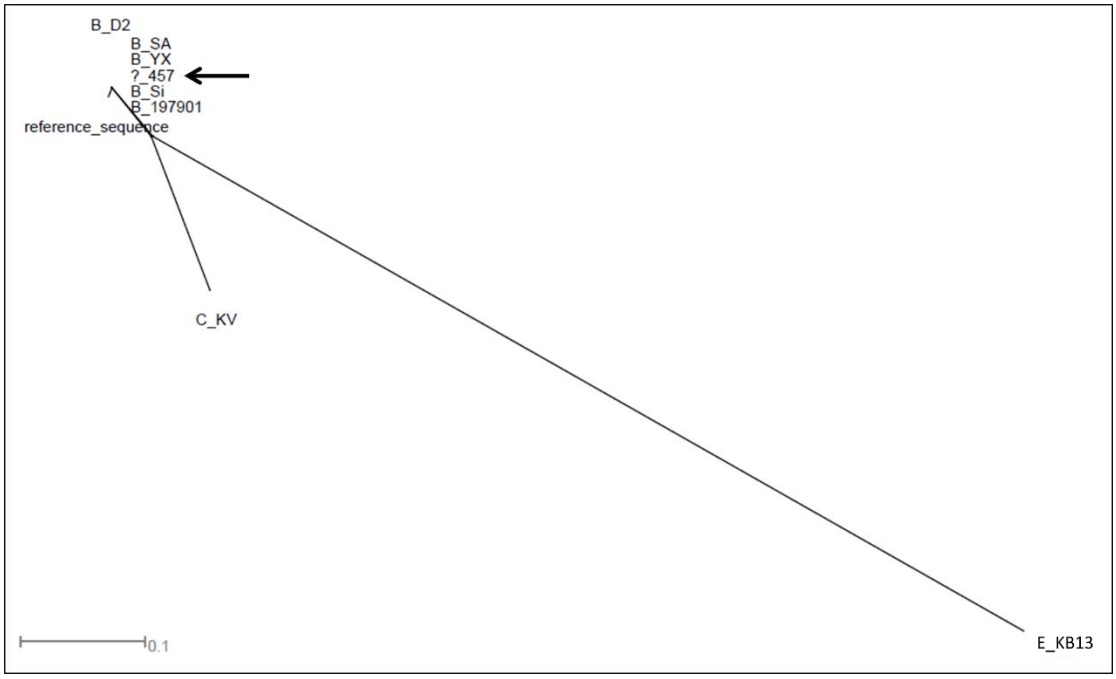


Figure 3.3 Phylogenetic network based on SNVs found in the whole genome of *A. astaci* belonging to genotypes B, C, E, and isolate 457. The distances between *A. astaci* isolates represent the genetic differences between isolates and genotypes. Reference sequence: genome assembly from *A. astaci* D2.

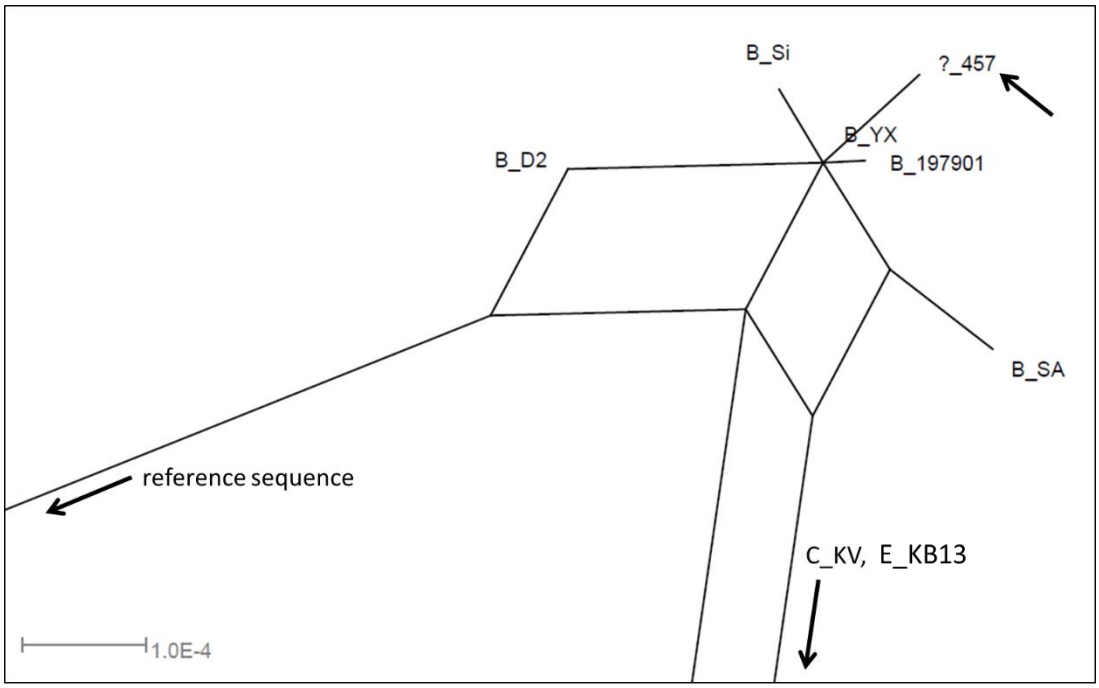


Figure 3.4 Magnification of phylogenetic network shown in Figure 3.3. The distances between *A. astaci* isolates represent the genetic differences between isolates and it is proportional to the significance of the split. The split network which intersects some of the edges allows additional significance to the split, showing more relationships between the isolates than what would be obtained by a phylogenetic tree. Reference sequence: genome assembly from *A. astaci* D2.

Phylogenetically, isolate 457 belongs with genotype B yet does not display a typical genotype-B RAPD-PCR profile. To investigate the basis of this apparent discrepancy, an *in silico* RAPD-PCR analysis with the primer B01 (Table 2.2) was carried out on *A. astaci* 457 and D2 genome assemblies. *A. astaci* isolate D2 was chosen as genotype B representative isolate, as it displays a typical RAPD-PCR profile. This analysis consisted of a *blastn* search of the primer sequence against the assembled genomes to retrieve sequences matching the sizes of RAPD-PCR bands on the gel. Unfortunately, it was not possible to match the sequences lengths as different sequence sizes were recovered from the search, probably due to the fragmented nature of the genome assemblies and the duration of the extension step in the PCR protocol. While the extension step time in a PCR protocol influences the size of the amplified PCR products, the genome assemble process is not perfect and gaps, misassembling of reads and missing contigs are factors that can influence the size of the products retrieved from the *blastn* search.

3.1.5 Cloning and sequencing of RAPD-PCR products from *Aphanomyces astaci* isolates 457 and D2 identify unique features.

Since the *in silico* analysis did not yield an explanation, another approach was therefore carried out to shed light on the aberrant RAPD-PCR profile of isolate 457: RAPD-PCR products from 457 (with aberrant profile) and from isolate D2 (with typical genotype-B profile) were excised from the agar gel after amplification and cloned in *E. coli*. Multiple colonies were screened (Figure 3.5 A) and clones sequenced. Several clones were sequenced for two PCR products corresponding to bands that distinguish the RAPD-PCR profiles of 457 and other genotype-B isolates. The primer B01 was detected and matched at the ends of the sequences of the cloned RAPD-PCR products, except for *A. astaci* D2-clone 1 which had a single-base mismatch on the reverse site. For each clone, consensus sequences were created with BioEdit and cut to start and finishing with B01 primer sequence (forward and reverse). Table 3.2 resumes the final lengths of the fragments.

Table 3.2 Lengths of *A. astaci* D2 and 457 RAPD-PCR cloned fragments. Lengths comprise the primers sites and were computed on the consensus sequences.

Isolate	Clone number	Fragment length (bp)	Coded name
A. astaci D2	1	862	D2_C1
	8	1208	D2_C8
A. astaci 457	1	1021	457_C1
	2	1831	457_C2
	3	1208	457_C3

The lengths of these fragments were compared to the sizes of RAPD-PCR products of the isolates as judged by their positions on the agarose gel. Figure 3.5 matches the cloned and sequenced products with bands on the RAPD-PCR profile gel: clones D2_C8 (pink *), 457_C1 (green *), 457_C2 (blue *) and 457_C3 (pink *) matched bands of similar sizes for all genotype B isolates with clones D2_C8 and 457_C3 sequences matching.

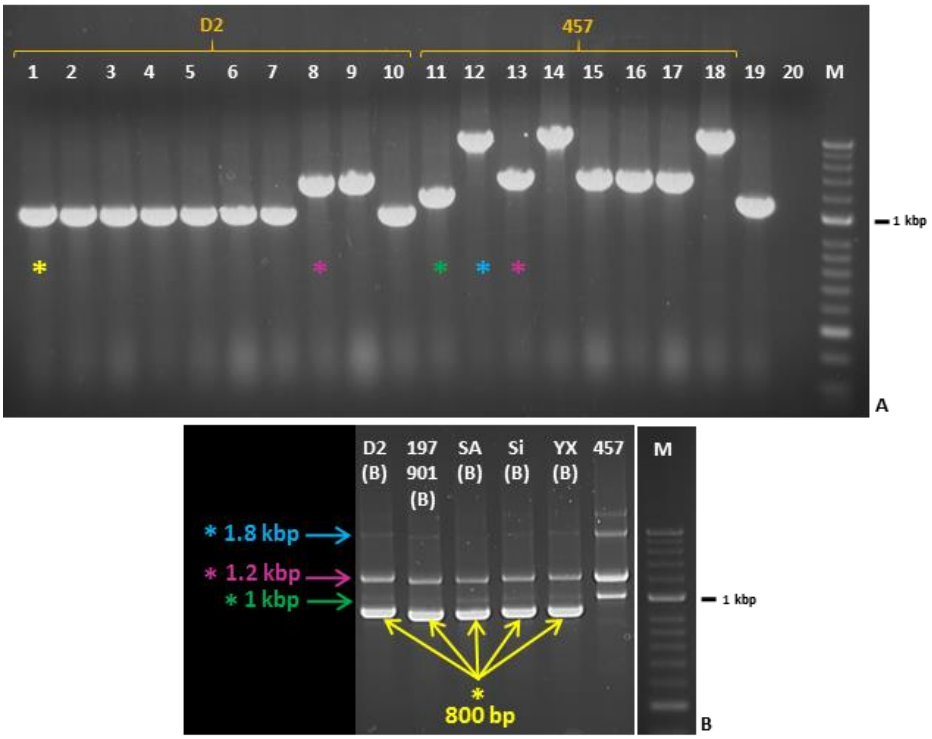


Figure 3.5 Comparison of cloned and transformed RAPD-PCR products from D2 and 457 isolates to genotype B RAPD-PCR profiles. **A**: colony PCR of *A. astaci* D2 (1-10) and 457 (11-18) cloned and transformed RAPD-PCR products; 19, positive control; 20, negative control. *, colour coded sequenced clones. **B**: RAPD-PCR profiles of *A. astaci* isolates belonging to genotype B; *, matching band to colour coded sequenced clones. M, Bioline HyperLadder II.

On the other hand, the band in D2 isolate matching the size of clone D2_C1 (Figure 3.5, yellow *), while present in profiles of all isolates belonging to genotype B, was not observed in the RAPD-PCR profile of isolate 457. Thus, this RAPD-PCR product is central to the aberrant RAPD-PCR profile seen in isolate 457 and was investigated further.

Using *blastn*, D2_C1 sequence was searched against the whole-genome assemblies of both isolates 457 and D2 and, for both, the first 519 bp (equal to 60.2 % of D2_C1 sequence) matched one contig with bit-score values above 770 and 94.03% of nucleotide sequence identity. Fragments of various lengths varying from 379 bp to 36 bp of the second half of D2_C1 (equal to 43.96 % to 4.17% of sequence) matched numerous other contigs with bit-score values lower than 500, and between 88.92 % to 100% of nucleotide sequence identity.

As the ~800 bp sequence was an important feature (bright band) for the genotype B isolates, the whole D2_C1 sequence was expected to be matching

at least one contig in D2 assembly. Thus, D2_C1 sequence was searched against the other isolates belonging to genotype B.

Blastn searches against genomes of isolates SA, Si, and YX presented similar outputs to the ones noted for D2 and 457 (first half of D2_C1 sequence matching one contig with high bit-score values, second half matching numerous contigs with lower bit-scores) while for isolate 197901 the *blastn* search found a single contig matching the whole length of D2_C1 (scf_5351_21236) and numerous contigs with low bit-scores matching the sequence second half (Figure 3.6).

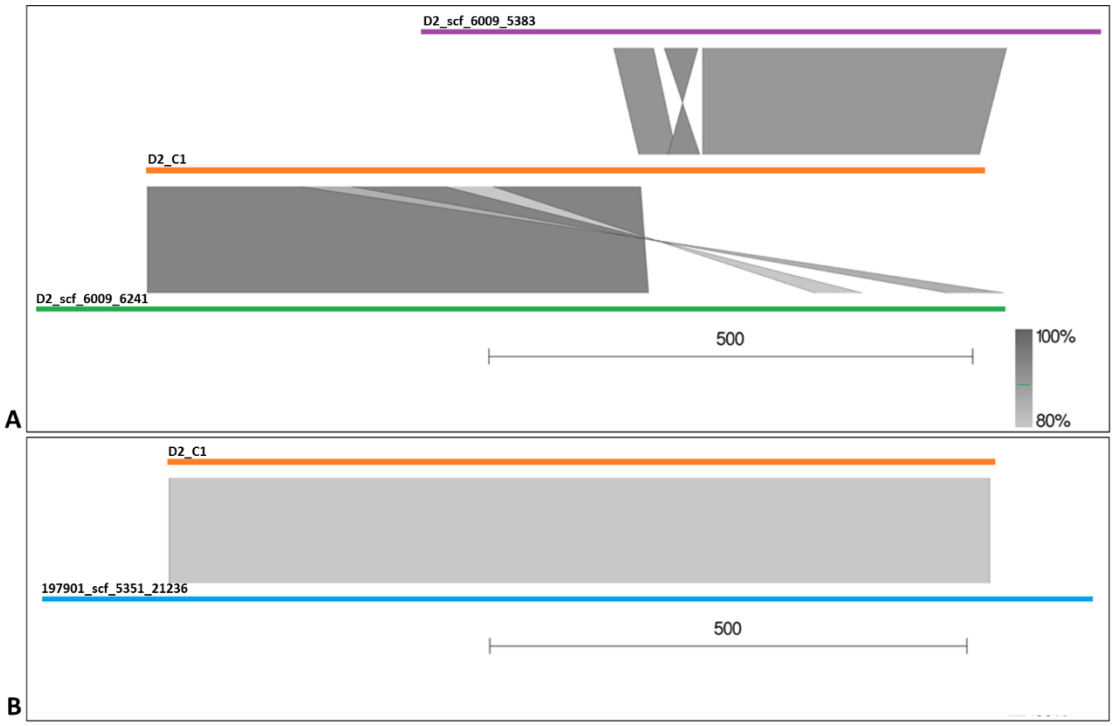


Figure 3.6 Graphic representation of selected alignments from BLASTN searches of D2_1C against D2 and 197901 assemblies. **A**: two of D2 contig (purple and green) matching two different segments of D2_C1 sequence (orange). **B**: 197901 contig (blue) matching the whole length of D2_C1 sequence (orange). In both figures: scale bar is 500 bp. Colour shaded grey diagrams of matching sequences vary in base of percentage of sequence identity.

This result confirmed that this RAPD-PCR product corresponded to a sequence that was present in the genome of at least one of the genotype B isolates. Probably, this sequence is present in the other genotype-B genomes too (except for isolate 457) and the failure to recover it intact from the genome assemblies reflects the fragmented nature of these draft genome assemblies. To verify this hypothesis, genomic shotgun sequencing reads from D2, SA, Si, YX and 457 were aligned to the 197901 assembly and the relevant section of

the alignment was visualised with IGV (Figure 3.7). Consistent with absence of the corresponding band in its RAPD-PCR profile, genomic reads from isolate 457 do not fully cover the genomic region shown in Figure 3.7. On the other hand, genomic reads from isolates D2, SA, Si and YX align against the whole length of the region with no breaks, supporting the presence of this sequence in the genomes of each of these isolates.

The correspondence of paired-end reads from D2, SA, Si and YX to the entire length of D2_C1 confirms the presence of the sequence in the genome of these isolates and is consistent with presence of a band of ~800 bp in their respective RAPD-PCR profile. This result strongly supports the hypothesis that this region of the genome has been misassembled in D2, SA, Si, and YX assembly. For more information about the assemblies, the reader is redirected to 3.6.

Besides this, 457 displays a substantial difference in coverage with paired-end reads covering only half of the contig and half of D2_C1 (Figure 3.7). This indicates that only half sequence of D2_C1 is present in the genome confirming the absence of a band of ~800 bp in 457 RAPD-PCR profile.

Figure 3.7 *A. astaci* genotype B isolates and 457 reads aligned to scf_5351_21236 (from isolate 197901), visualised in IGV. Reference genome: *A. astaci* 197901. From top to bottom paired-end reads of: *A. astaci* 457, D2, SA, Si, YX, and 197901 (genotype B). Red bar in navigation panel: D2_C1 whole sequence. Vertical bars: depth of read coverage for each genomic position against the reference genome. Grey bars: majority (> 80 %) of nucleobases aligned match the reference genome. Green, red, yellow, and blue bars: at least 20 % of nucleobases aligned not matching the reference genome and substituted respectively by adenosine (A), tyrosine (T), guanosine (G), and cytosine (C). Reads colour coding scheme: IGV default paired end reads colouring scheme; reads are flagged and coded by the chromosome/scaffold on which their mate pair can be found.

Visually checking (using IGV) parts of the genome with similarity to RAPD-PCR products revealed heterozygosity at some of these loci in all the genotype-B isolates except for isolate 457. Elsewhere in the genome, isolate 457 showed similar patterns of heterozygosity to the other isolates. This suggests that isolate 457 may have undergone localised loss of heterozygosity (LOH) (Figure 3.8). In this figure, isolate 457 reads are at the bottom of the list and each nucleobase aligned to the reference assembly is grey, indicating that those positions are homozygous. However, aligned reads from the other isolates belonging to genotype B (*i.e.* 197901, D2, SA, Si, and YX) display variability on the content of nucleobases and various nucleobases are indeed heterozygous (Figure 3.8 A). LOH for isolate 457 was present not only in the section of the contig matching the D2_C1 clone, but the whole contig presented various degree of homozygosity.

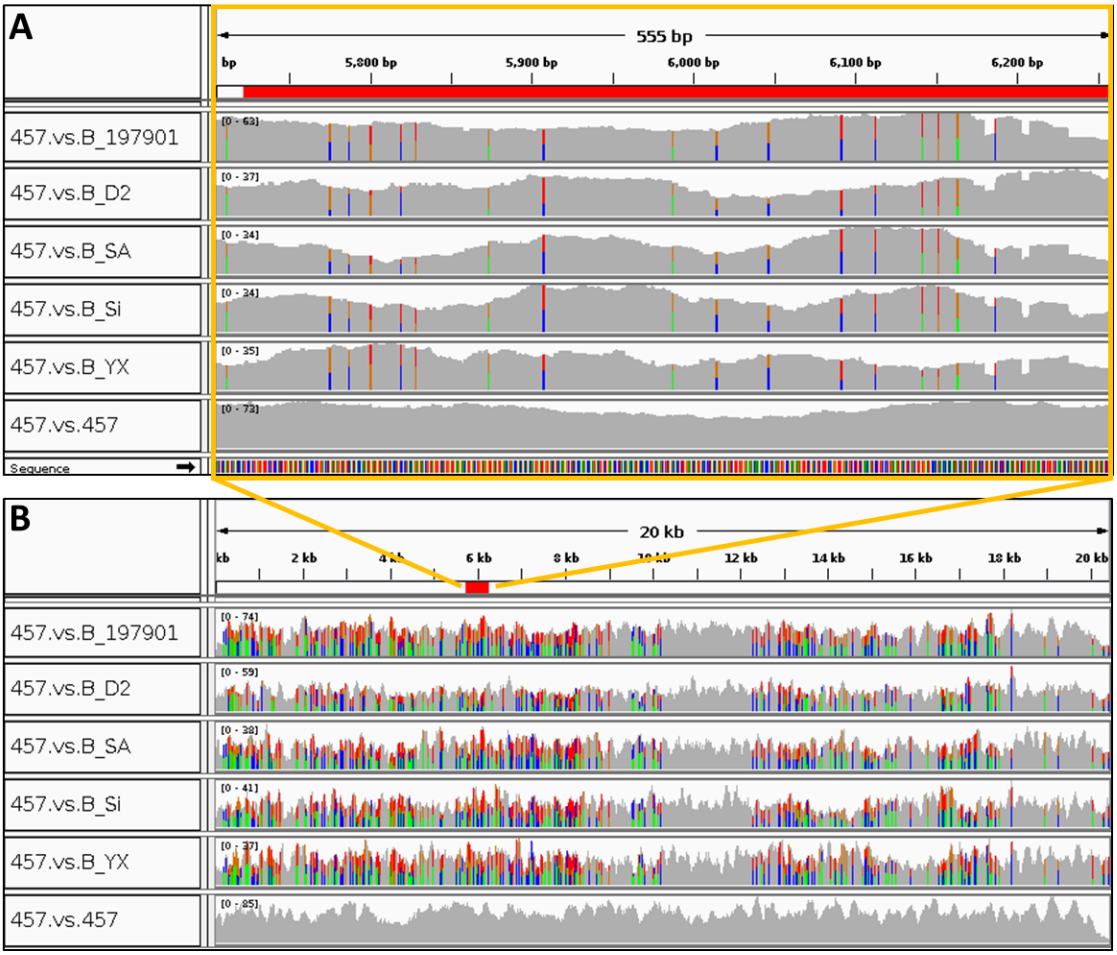


Figure 3.8 *A. astaci* 457 loss of heterozygosity visulised in IGV. Both figures, from top to bottom, paired-end reads of: *A. astaci* 197901, D2, SA, Si, YX (genotype B) and 457. Reference genome: *A. astaci* 457. Red bar in navigation panel: part of D2_C1 sequence matching 457 contig. **A**: section of 457 contig matching D2_C1 displaying loss of heterozygosity; **B**: whole 457 contig displaying loss of heterozygosity. Vertical bars: depth of read coverage for each genomic position against the reference genome. Grey bars: majority (> 80 %) of nucleobases aligned match the reference genome. Green, red, yellow, and blue bars: at least 20 % of nucleobases aligned not matching the reference genome and substituted respectively by adenosine (A), tyrosine (T), guanosine (G) and cytosine (C).

3.1.6 Discussion

As previously mentioned (section 1.2), *A. astaci* is a parasite of freshwater crayfish, involved in the decline of European indigenous crayfish species (Alderman *et al.*, 1984; Unestam, 1972). Five genotypes (A, B, C, D, and E) have been detected and genetically characterised from pure cultures by RAPD-PCR, the preferred method to genotype this parasite (Diéguez-Uribeondo *et al.*, 1995; Huang *et al.*, 1994; Kozubíková *et al.*, 2011). Other fingerprinting tools developed include AFLP and microsatellites typing which noted variation in genotypes as defined by RAPD-PCR (Grandjean *et al.*, 2014; Rezinciuc *et al.*, 2014). Epidemiological studies of crayfish plague dependent on the distinction of *A. astaci* genotypes proved to be a useful tool to better understand the history and spread of this disease in Europe (Lilley *et al.*, 1997a; Rezinciuc *et al.*, 2014; Viljamaa-Dirks *et al.*, 2013; Vrålstad *et al.*, 2014) and being able to identify quickly and reliably *A. astaci* genotypes is an important step to study this disease.

In the present study, the genotypes of the *A. astaci* isolates held in the OCC were checked by RAPD-PCR and all but one could be unambiguously assigned to one of the previously described genotypes. The investigation of the atypical *A. astaci* isolate 457 RAPD-PCR profile by SNVs and cloned RAPD-PCR products established the affinity of this isolate to genotype B and the identification of localised LOH in loci targeted by the RAPD-PCR that directly impacted the RAPD-PCR profile: phylogenetically, isolate 457 belongs within genotype B but because of a localised LOH at loci targeted by the RAPD-PCR assay, this assay yields an aberrant result. To clarify the position of isolate 457 in the genotypes classification several thousand of SNVs were identified for each known genotype and used to construct phylogenetic networks (section 3.1.4). In the network construction (Figure 3.2, Figure 3.3 and Figure 3.4), it is apparent that the genetic differences separating genotypes A and D from B, C and E are numerous. These differences could reflect the history and the host-parasite interaction. In particular, the genetic differences that establish the divergence of genotype D from other genotypes could be related to its adaptation at higher temperatures, in contrast with other genotypes which are adapted to lower temperatures (Rezinciuc *et al.*, 2014). The genetic differences that establish the divergence of genotype A, instead, have not been well

documented yet as this genotype have been detected only from European crayfish species while the detection from the original North American crayfish host is still missing. Isolates belonging to genotype A are thought to be related to the first introduction of *A. astaci* in Europe and infection studies on genotypes A and B indicate differences in virulence between these genotypes and between isolates belonging to genotype A (Becking *et al.*, 2015; Makkonen *et al.*, 2014; Makkonen, *et al.*, 2012; Viljamaa-Dirks *et al.*, 2016). These virulence differences could be due to genetic variants that separate this genotype from the others. Moreover, even though genotypes B, C, and E grouped close together, the SNVs between them are sufficient to unambiguously distinguish these three genotypes, concordant with the genetic differences identified by RAPD-PCR. Considering this approach, it can be concluded that isolate 457 belongs to genotype B. This genotype was firstly isolated from *P. leniusculus* (signal crayfish) and from European indigenous freshwater crayfish after the introduction of *P. leniusculus* in Europe; genotype C is based on one single isolate from *P. leniusculus* native from Canada and imported into Sweden; genotype E was first isolated from *O. limosus* (spiny-cheek crayfish) in the Czech Republic and the most recent one to be characterised (Diéguez-Uribeondo *et al.*, 1995; Huang *et al.*, 1994; Kozubíková, *et al.*, 2011; Lilley *et al.*, 1997a; Rezinciuc *et al.*, 2014).

The Perl script used to obtain SNVs information to create the phylogenetic analysis, uses data regarding each base position of aligned reads against a reference genome. To acquire only single nucleotides sites which represented a true variant, the script parameters were set stringent enough to avoid heterozygous sites (see section 2.4.8 for more details). Thus, information about heterozygous sites and unique features (*i.e*. gaps, insertions, unique sequences) from both the reference genome and the reads aligned are not considered. Hence, the genomic differences noted in the RAPD-PCR profile of isolate 457 could be located in these parts of genome that are ignored by the script. The atypical 457 RAPD-PCR profile was therefore investigated by direct analysis of RAPD-PCR bands (section 3.1.5). Cloning the RAPD-PCR bands from isolate 457 and D2 (the latter exhibiting a typical genotype B RAPD-PCR profile and therefore chosen to represent genotype B) and comparing the cloned sequences to the genotype B isolates genomes allowed the identification

of unique features in isolate 457. In particular, by investigating an important bright band (clone D2_C1) present the RAPD-PCR typical profile for genotype B isolates but missing in isolate 457, the absence of a portion of genome and LOH were located in isolate 457 genome (Figure 3.7 and Figure 3.8). These two features characteristic of isolate 457 are in sites typically targeted by the B01 primer used in the RAPD-PCR protocol and explain isolate's 457 atypical RAPD-PCR profile.

LOH has been previously observed in oomycetes pure cultures kept in laboratories for a long time and from fresh isolates collected in the wild (Dobrowolski *et al.*, 2002; Hu *et al.*, 2013; Lamour *et al.*, 2012; Shrestha *et al.*, 2014). This characteristic has been associated in particular with changing in mating strategies during asexual growth, typical situation that occurs in laboratory condition (Shrestha *et al.*, 2014). The exact meaning of LOH and why it occurs is still not clear in oomycetes , but it has been suggested that LOH can reduce and remove dominant alleles and traits to favour the recessive alleles that could contribute towards adaptation to new environments (Jiang *et al.*, 2013; Kasuga *et al.*, 2016). LOH has been associated also with loss of pathogenicity (Lamour *et al.*, 2012), a characteristic that could be an advantage for *A. astaci* to successfully coexists with the European crayfish hosts (Jussila *et al.*, 2016). The loss of pathogenitcity and virulence of *A. astaci* isolates has been experimentally proved by numerous trials (Aydin *et al.*, 2014; Jussila *et al.*, 2013, 2016; Makkonen *et al.*, 2012; Viljamaa-Dirks *et al.*, 2016) and the hypothesis of the presence of adaptations to European crayfish has been supported by the detection of latent infections(Jussila *et al.*, 2011; Kušar *et al.*, 2013; Viljamaa-Dirks *et al.*, 2011) which could be a result of LOH, but until now this hypothesis has not been verified.

While RAPD-PCR allows the distinction of all known *A. astaci* genotypes, it presents some downsides that can influence the reproducibility and repeatability of the test. The difficulties in interpreting RAPD-PCR results have been investigated by numerous authors (Jones *et al.*, 1997; Penner *et al.*, 1993; Skroch and Nienhuis, 1995) whose studies show that the reproducibility of RAPD-PCR test highly depends on the standardisation of PCR reaction conditions. As the RAPD method is a PCR based technique, it is also subjected to the same complications that one can incur doing a standard PCR. For

example, the variability of the PCR machine and reagents used, the sampling techniques and the concentration of DNA template, are all factors that influence the reproducibility of the test in different laboratories on the same organisms (Jones *et al.*, 1997; Penner *et al.*, 1993; Skroch and Nienhuis, 1995). In addition, the low stringency of homology of the primer used in the RAPD-PCR constitutes a competition effect during the amplification process that can influence the result itself (Halldén *et al.*, 1996; Penner *et al.*, 1993). Thus, the correct interpretation of RAPD-PCR results can be challenging and it has been suggested that various primers and markers should be screened to define a specific genotype, selecting the ones with optimal characteristics and thus making this approach informative (Wilkerson *et al.*, 1993; Williams *et al.*, 1990). As suggested by the name of the technique, this method randomly amplifies DNA regions that are not necessarily known to the user (Ali *et al.*, 2004). While this can be an advantage to detect large quantities of polymorphisms, it can also be considered a disadvantage as the markers (bands on the electrophoresis gel) migrate by size (quantitative discrimination) and it is not possible to differentiate them by sequence (qualitative discrimination) (Kumari and Thakur, 2014), meaning that bands of same size can correspond to different DNA fragments/regions (Harris, 1999). Another disadvantage is that all markers obtained by the application of this technique are considered dominant and the profile of a genotype is defined by the presence/absence of multiple amplified DNA segments of unknown sequence (Ali *et al.*, 2004; Kumari and Thakur, 2014; Williams *et al.*, 1990). Moreover, it is not possible to distinguish if the amplified segment is homozygous or heterozygous. The presence of a band on the agarose gel after PCR amplification can results by the complete alignment of the primer to the DNA template but also results from a partial alignment. The absence of a band can results by the presence of a true polymorphisms in the primer site or can also be due to the competitive nature of PCR, template-primer misalignment or degraded DNA (Ali *et al.*, 2004; Kumari and Thakur, 2014; Tingey and del Tufo, 1993; Williams *et al.*, 1990). *A. astaci* isolate 457 presents genetic features that influenced its RAPD-PCR profile normally targeted by the RAPD-PCR primer B01 and displayed by genotype B isolates. Thus, the precise interpretation of RAPD-PCR results can be challenging and other markers are necessary. In this study, *A. astaci* 457 has been included in the genotype B isolates, following the SNVs analysis.

Moreover the LOH observed in *A. astaci* 457 might have occurred during the sub-culturing of this isolate in laboratories, causing the modification of its RAPD-PCR profile originally described by Lilley *et al.* (1997a).

3.2 Investigation on the genotyping potential of widely used molecular markers

The aims of the present study were to verify the identity of *Aphanomyces* spp. isolates held in the Cefas OCC, to check for genetic heterogeneity within each isolate, to determine their phylogenetic positions and to check if widely used molecular markers could discriminate between *A. astaci* genotypes as defined by RAPD-PCR (Diéguez-Uribeondo *et al.*, 1995; Huang *et al.*, 1994; Kozubíková *et al.*, 2011). These aims were addressed using genetic markers previously described in the literature: the internal transcribed spacer of the rDNA (ITS), the D1-D2 regions of the large subunit ribosomal DNA (LSU) and the cytochrome oxidase subunit I (COI). These markers were chosen to explore their ability to distinguish genera, species and isolates belonging to the same species.

In particular, ITS regions of the ribosomal DNA are a useful marker to separate taxa at different levels of resolution, ranging from genus to species level. ITS is the most employed DNA barcode used by the mycology community to identify and discriminate between species (Bruns *et al.*, 1991; Lee and Taylor, 1992; Robideau *et al.*, 2011; Schoch *et al.*, 2012) and it is also the preferred marker for oomycetes (Cooke *et al.*, 2000; Lévesque and De Cock, 2004; Robideau *et al.*, 2011). It has been recently used to elucidate the phylogenetic relationship between *Aphanomyces* species (Diéguez-Uribeondo *et al.*, 2009).

The LSU region of the ribosomal DNA has been used to construct phylogenetic analyses in fungi (Porter and Golding, 2012), yeast (Fell *et al.*, 2000) and oomycetes (Leclerc *et al.*, 2000; Petersen and Rosendahl, 2000; Riethmüller *et al.*, 1999; Robideau *et al.*, 2011). This phylogenetic marker is characterised by divergent domains, known as “D” regions, containing variable nucleotide sequences. In oomycetes, this marker is mostly used at family level (Leclerc *et al.*, 2000; Petersen and Rosendahl, 2000) but it also gives resolution at species level for the genera *Achlya* (Leclerc *et al.*, 2000) and *Pythium* (Lévesque and De Cock, 2004).

COI is a mitochondrial gene located on the mtDNA and adopted as the official barcode by the Consortium for the Barcode of Life (www.barcodeoflife.org) and

by GenBank (http://www.ncbi.nlm.nih.gov/WebSub/?tool=barcode). Due to its high rate of base substitutions, it is an effective phylogenetic/taxonomic marker for eukaryotes species and it can separate both closely related species and isolates within the same species (Hebert *et al.*, 2003; Schindel and Miller, 2005)(Hebert *et al.*, 2003; Schindel and Miller, 2005). This barcode has been successfully applied to oomycetes, proving that this molecular marker can be a useful tool to discriminate closely related species in this class (Bala *et al.*, 2010; Martin and Tooley, 2003; Robideau *et al.*, 2011).

3.2.1 Aims and objectives

In this chapter, ITS, COI and LSU molecular markers were amplified from all the *Aphanomyces* isolates held in the Cefas OCC and cloned. Clones were sequenced to check genetic heterogeneity within each isolate (intragenomic variation) and between isolates belonging to the same species (intraspecific variation). For each molecular marker, phylogenetic trees were constructed to define the position of the isolates used in the study and to check the markers could distinguish between *A. astaci* genotypes.

3.2.2 Materials and methods

All methods and materials are described in detail in Chapter 2 – Materials and Methods. Briefly, gDNA was extracted from the 20 isolates held in the OCC (Table 2.1) as described in section 2.3.2. Genomic DNA was amplified by PCR (section 2.3.4.1 for ITS, 2.3.4.3 for COI and 2.3.4.4 for LSU fragments). PCR fragments were checked by gel electrophoresis (section 2.3.1), purified (section 2.3.5), cloned into pJET1.2 (Thermo Scientific) or into pGEM-T-Easy (Promega) vectors (section 2.3.6) and transformed into *E. coli* DH5α competent cells (section 2.2.3). Colonies were screened by PCR, positive plasmids extracted with QIAprep Spin Miniprep Kit (Qiagen, Germany) (section 2.3.7) and between one to six clones for each isolate were sequenced to detect variability in the COI, ITS, and LSU regions (section 2.3.8). Sequences were visualised with BioEdit version 7.0.8 (Hall, 1999) and checked with BLAST against the NCBI Nucleotide database (section 2.4.1). Internal primers for the LSU fragments

were designed to cover the whole sequence length: primer IF was used for *A. astaci* and *A. invadans* isolates; primer 95 for isolate *A. invadans*-like NJM9510; primer FRIG for *A. frigidophilus* isolates; primer IR for all isolates (Table 2.2). For COI and ITS genetic markers, the whole sequence length was covered in one sequence read. Sequences were aligned with MUSCLE (Edgar, 2004) and viewed with MEGA6 software (Tamura *et al.*, 2013) (section 2.4.1 and 2.4.2). Sequence differences between clones and isolates were noted, consensus sequences were generated with BioEdit (section 2.4.1). For each genetic marker, phylogenetic analyses were carried out as per section 2.4.2.

3.2.3 Analysis of the internal transcribed spacer (ITS) confirms identities of working cultures and confirms that isolate NJM9510 belongs to a distinct, as-yet undescribed, species

In the present study, the ITS region of each *Aphanomyces* isolate (Table 2.1) was amplified with primers ITS5 and ITS4 (Table 2.2). The PCR products were cloned and sequenced. PCR-products lengths for each species were obtained by aligning the sequenced clones and include the primers sites (Table 3.3).

Table 3.3 Lengths (bp) of ITS fragments for the *Aphanomyces* isolates used in this study. Lengths listed by species.

Species	Length of ITS fragment (bp)
A. astaci	777
A. frigidophilus	765
A. invadans	764
"A. invadans-like" NJM9510	781

Intragenomic single nucleotide variants were detected among sequences of PCR products cloned from a single isolate. For *A. astaci*, out of 777 sites, 40 were variable among the 57 cloned ITS sequences with p=5.1 %, in which p represents the percentage of variants in relation to the fragment length. For *A. frigidophilus*, out of 765 sites, 10 (p=1.3 %) were variable among 16 cloned sequences. For *A. invadans*, out of 764 sites, 14 (p=1.8 %) were variable among 23 cloned sequence. For *A. invadans*-like NJM9510, out of 781 sites, 7 (p=0.9%) were variable among 5 cloned sequences. For a complete list of nucleotide variants see Table 5.8. Consensus sequences for each isolate were

generated with BioEdit and BLAST searched against previously available homologous sequences in the NCBI Nucleotide database. These searches confirmed the expected identities of the isolates as *A. astaci*, *A. frigidophilus* and *A. invadans*. The single exception was isolate *A. invadans*-like NJM9510, whose 781 bp ITS sequence matched that of *Aphanomyces* sp. 84-1240 (GenBank accession number AF396683) with 99 % identity. The sequences retrieved from NCBI Nucleotide database were used to generate an ITS-based phylogenetic tree and isolate NJM9510 fell within the lineage containing saprotrophic/opportunistic species and not within *A. invadans* (Figure 3.9).

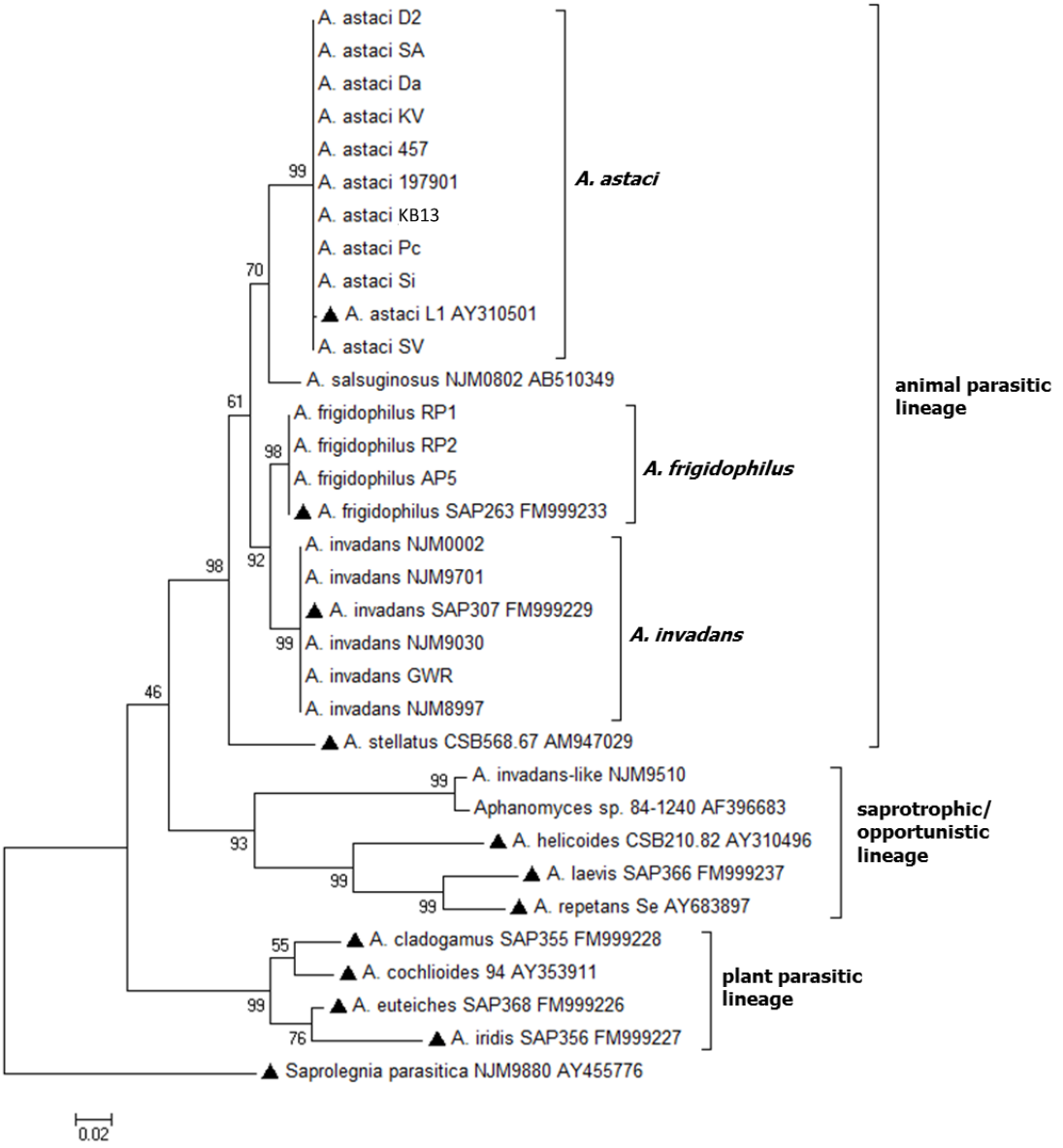


Figure 3.9 Molecular phylogenetic analysis of ITS data matrix. Matrix build with sequences from isolates held in the OCC and sequences retrieved from BLAST searches (Table 2.1 and Table 5.5) and analysed by maximum likelihood, based on the Tamura-Nei model (Tamura and Nei, 1993). Numbers adjacent to each branch represent the percentage of bootstrap support calculated for 500 replicates. *Saprolegnia parasitica* was used as outgroup. Species, isolate ID and GenBank reference number are indicated. ▲ Reference sequences (Diéguez-Uribeondo *et al.*, 2009).

Intraspecific variants were detected when comparing our consensus sequences data to the ones retrieved from the NCBI Nucleotide database. For *A. astaci*, the percentage identity of the sequences ranged between 97 and 100 % and between 99 and 100 % for *A. invadans* and *A. frigidophilus*, respectively. The ITS sequence from "*A. invadans*-like" NJM9510 shared 99% identity with *Aphanomyces* sp. 84-1240 (GenBank accession number AF396683.1) and less than 83 % with other *Aphanomyces* species (*A. helicoides*, *A. laevis* and *A. repetans*). The observation of intraspecific genetic variation raised the question

of whether these ITS differences can separate *A. astaci* isolates in genotypes and if these groups corresponds with the previously defined RAPD-based genotypes. A separate phylogenetic analysis was performed to reveal if the intraspecific nucleotide variants noted between *A. astaci* isolates could ascribe these and the *A. astaci* sequences retrieved from NCBI Nucleotide database to their genotypes (Figure 3.10).

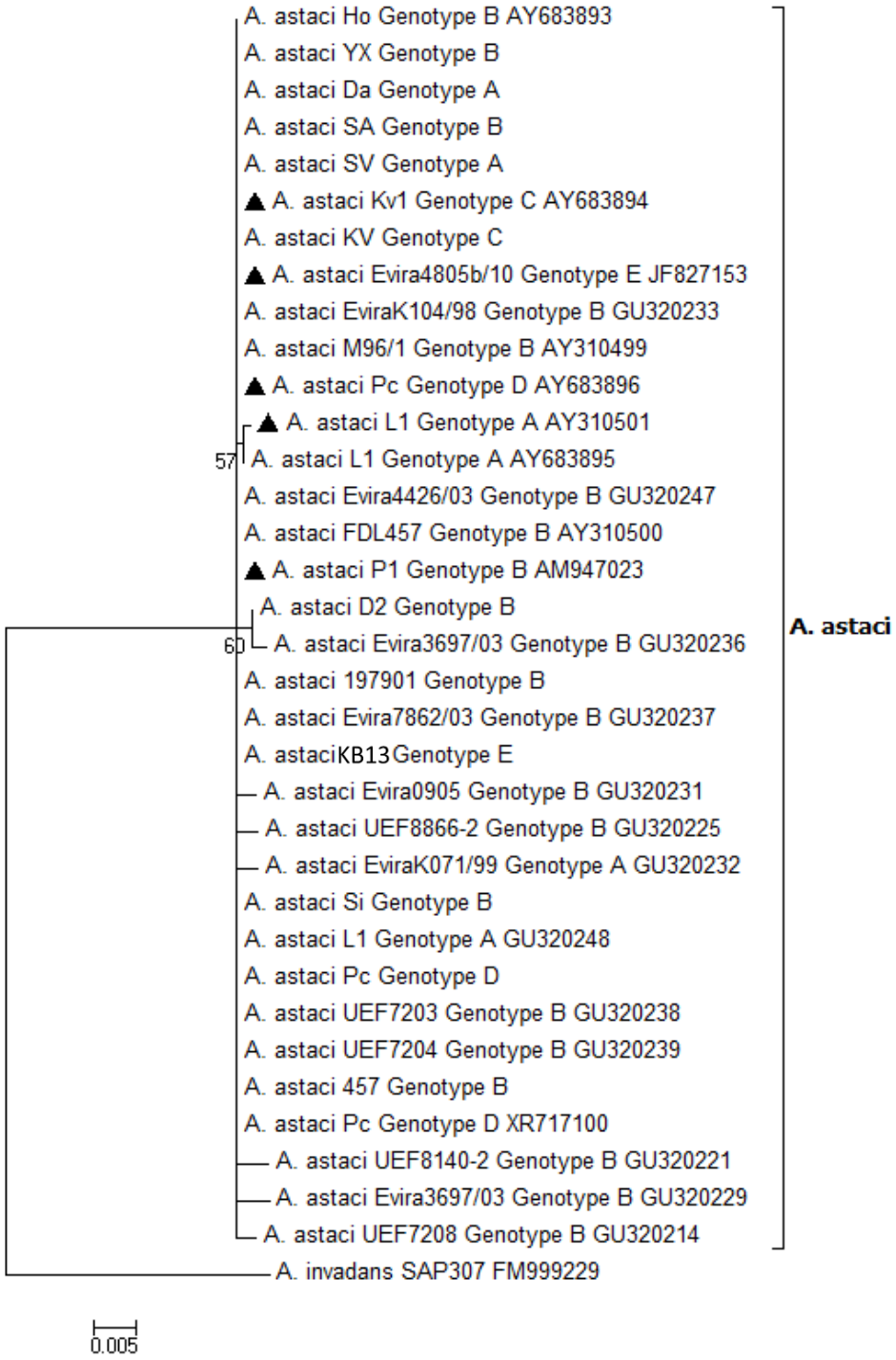


Figure 3.10 Molecular phylogenetic analysis of ITS data matrix from *A. astaci* isolates. Matrix build with sequences from isolates held in the OCC and sequences retrieved from BLAST searches (Table 2.1 and Table 5.6) and analysed by maximum likelihood, based on the Tamura-Nei model (Tamura and Nei, 1993). Numbers adjacent to each branch represent the percentage bootstrap support calculated for 500 replicates. *A. invadans* was used as outgroup. Species, isolate ID, genotype and GenBank reference number are indicated. ▲ *A. astaci* genotype-reference sequences (Table 5.6).

In this study, the intraspecific nucleotide variants noted between *A. astaci* isolates were not consistent within genotypes and could not discriminate one genotype from the other (Figure 3.10 and Figure 5.1).

3.2.4 The phylogeny of the D1-D2 regions of the large subunit ribosomal DNA (LSU) is consistent with the ITS phylogeny

Since the ITS fragment proved to be unsuitable to distinguish among *A. astaci* intraspecies genotypes, the potential of the LSU fragment was investigated. This region shares many of the advantages of the ITS but is significantly longer and therefore might hold more phylogenetic signal. The sequence considered contains partial 18 S rDNA gene, ITS1 region, 5.8 S rDNA gene, ITS2 region and partial 28 S rDNA gene.

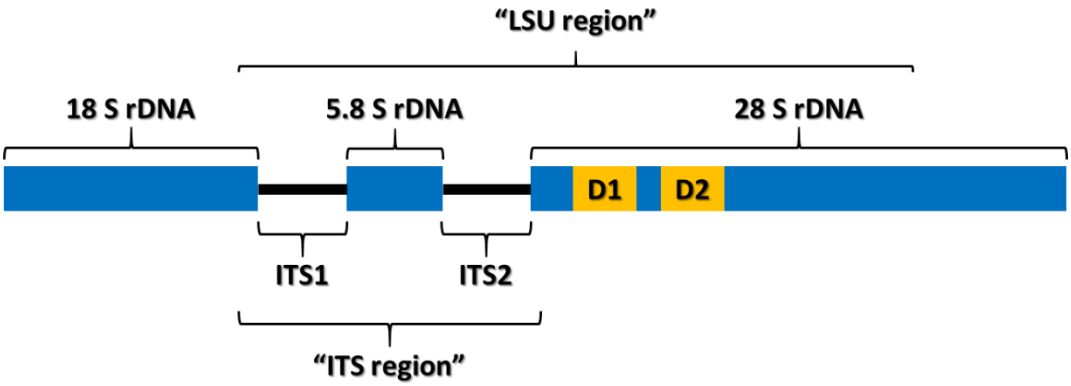


Figure 3.11 Schematic representation of "ITS" and "LSU regions" considered in this study.

The LSU region of the ribosomal DNA of each *Aphanomyces* was amplified, cloned, and sequenced using primers UNup18S42 and UN-lo28S1220 (Table 2.2). PCR-product lengths for each species were obtained by aligning the sequenced clones and include the primers sites (Table 3.4)

Table 3.4 List of lengths (bp) of the LSU fragments, divided by species and including the primers sites.

Species	Length of LSU fragment (bp)
A. astaci	2108
A. frigidophilus	2096
A. invadans	2095
"A. invadans-like" NJM9510	2116

The availability of LSU sequences from animal pathogenic *Aphanomyces* in the NCBI Nucleotide database is limited, as ITS has been the most used marker for phylogenetic analysis in oomycetes and the most targeted region for detection of *Aphanomyces* (Kozubíková *et al.*, 2009; Makkonen *et al.*, 2011; Oidtmann *et al.*, 2004; Oidtmann *et al.*, 2006; Robideau *et al.*, 2011).

Comparison of *A. astaci*, *A. frigidophilus* and *A. invadans* consensus sequences generated in this study against the NCBI Nucleotide database, resulted in alignments with *Aphanomyces* species with identity values ranging between 93 to 100 %. For *A. invadans*-like NJM9510 the highest identity value was 96 % with *A. laevis* (GenBank accession number HQ665242.1) and *Aphanomyces* sp. (GenBank accession number HQ665276.1). The sequences retrieved from the NCBI Nucleotide database that covered the whole length of the LSU fragment were limited, thus only few sequences were compared with the consensus sequences generated in this study for the phylogenetic analysis (Figure 3.12).

As for the ITS marker, intragenomic single nucleotide differences were detected (Table 5.9). For *A. astaci*, out of 2108 sites, 24 (p=1.1 %) were variable among the 17 cloned sequences. For *A. frigidophilus*, out of 2096 sites, 4 (p=0.2 %) were variable among 10 cloned sequences. For *A. invadans*, out of 2095 sites, 11 (p=0.5 %) were variable among the 17 cloned sequences. For NJM9510, out of 2116 sites, one (p=0.05 %) was variable among the 3 cloned sequences. Consensus sequences for each isolate were generated with BioEdit, checked against the NCBI Nucleotide database, and used for the phylogenetic reconstruction.

The phylogenetic analysis based on the LSU marker (Figure 3.12) was consistent with the ITS-based analysis in which *A. invadans*-like NJM9510 grouped within the saprotrophic/opportunistic species, a lineage previously described by Diéguez-Uribeondo *et al.* (2009). The other isolates grouped within their respective species (Figure 3.12).

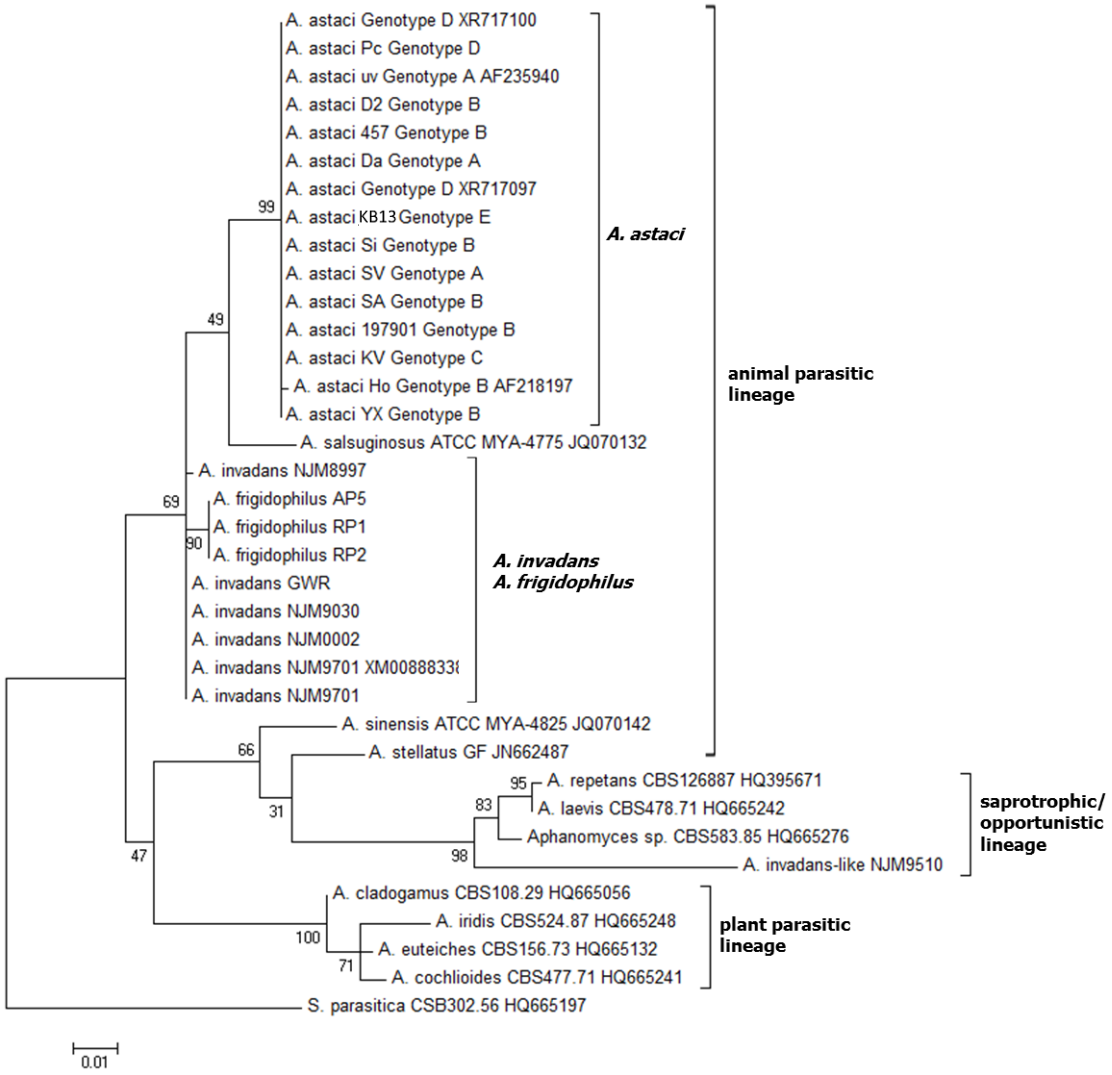


Figure 3.12 Molecular phylogenetic analysis of LSU data matrix. Matrix build with sequences of isolates held in the OCC and sequences retrieved from BLAST searches (Table 2.1 and Table 5.7) analysed by maximum likelihood based on the Tamura-Nei model (Tamura and Nei, 1993). Numbers adjacent to each branch represent the percentage of bootstrap support calculated for 500 replicates. *Saprolegnia parasitica* was used as outgroup. Species, isolate ID, genotype and GenBank reference number are indicated. Lineages follow Diéguez-Uribeondo *et al.* (2009).

3.2.5 Cytochrome oxidase subunit I (COI) based phylogeny is consistent with ITS and LSU analysis and discriminate *A. astaci* genotype D

As ITS and LSU markers were not able to discriminate between *A. astaci* genotypes, COI fragments were amplified, cloned, and sequenced to test whether they would provide additional phylogenetic signal. As ITS has been the preferred marker for phylogenetic analysis in oomycetes, only a few published studies include COI as genetic marker (Robideau *et al.*, 2011) and, consequently, the availability of *Aphanomyces* spp. COI sequences in public databases is limited.

For all isolates, the length of the COI PCR-amplified fragments was obtained by aligning the sequenced clones and was 751 bp including primers sites. As for ITS and LSU sequences analysed in this study, intragenomic nucleotide variants were present (Table 5.10). For *A. astaci*, out of 751 sites, 25 (p=3.3 %) were variable among the 47 cloned sequences. For *A. frigidophilus*, out of 751 sites, 5 (p=0.7 %) were variable among the 16 cloned sequences. For *A. invadans*, out of 751 sites, 13 (p=1.7 %) were variable among 24 cloned sequences. For NJM9510, out of 751 sites, one (p=0.1 %) was variable among the 5 cloned sequences. For *A. astaci*, *A. frigidophilus* and *A. invadans* isolates, performing BLAST searches against the NCBI Nucleotide database resulted in matches against *Aphanomyces*, *Achlya* and *Saprolegnia* species, with the highest identity values ranging between 91 to 93 %. For *A. invadans*-like NJM9510 the highest identity was with *A. laevis* (GenBank accession number HQ708195) with identity value of 96 %. Consensus sequences for each isolate were then generated and no intraspecific variants were detected in *A. frigidophilus* and *A. invadans*. On the other hand, in *A. astaci* isolate Pc (genotype D) presented significant intraspecific variants when compared to the other isolates (Figure 5.2). A total of three SNVs were counted in the consensus sequence generated from *A. astaci* isolate Pc (genotype D).The phylogenetic tree generated with this marker (Figure 3.13) has a similar partition of isolates and species as the ITS and LSU trees generated in this study and the one previously described by Diéguez-Uribeondo *et al.* (2009), with *A. invadans*-like NJM9510 grouping within the saprotrophic/opportunistic lineage, confirming once again that this isolate does not belong to *A. invadans* species.

A. astaci isolate Pc (genotype D) grouped within the other *A. astaci* isolates, but separately from them (Figure 3.13).

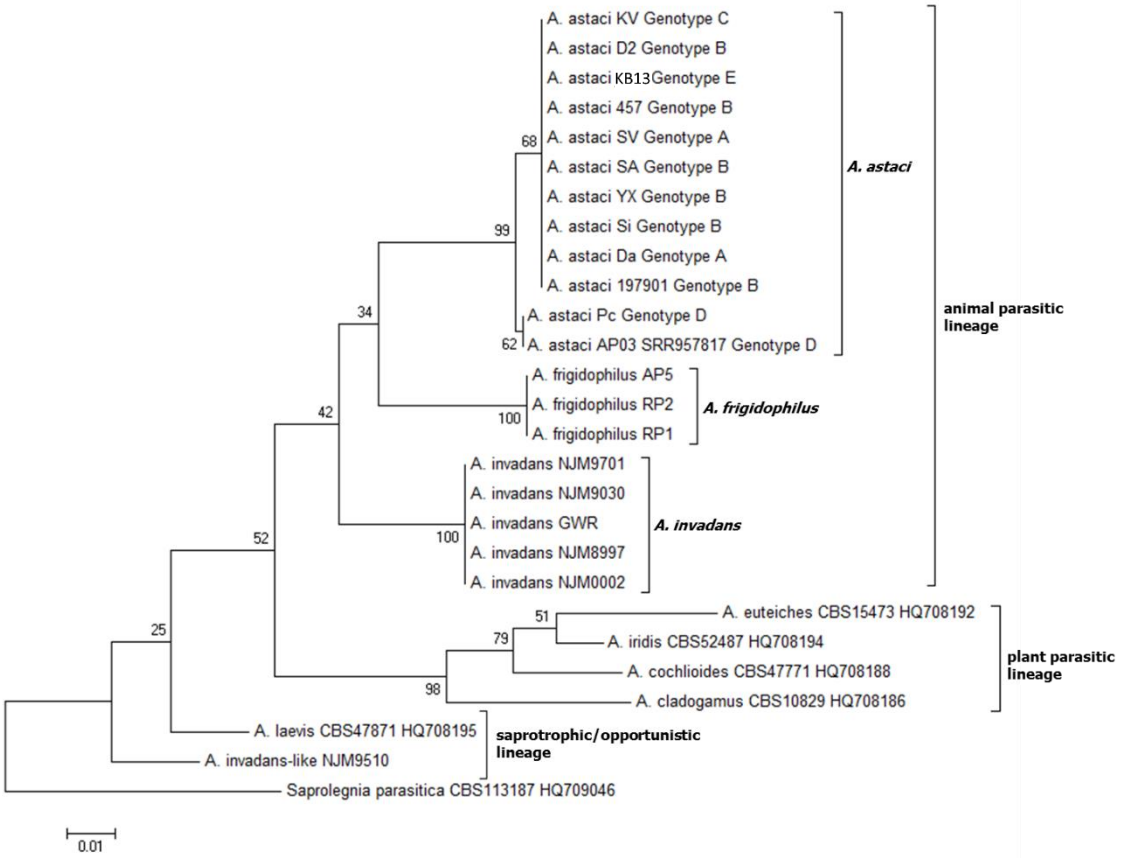


Figure 3.13 Molecular phylogenetic analysis of COI data matrix. Matrix build with sequences from the isolates held in the OCC and the sequences retrieved from the BLAST search (Table 2.1 and Table 5.4) and analysed by maximum likelihood based on the Tamura-Nei model (Tamura and Nei, 1993). Numbers adjacent to each branch represent the percentage bootstrap support calculated for 500 replicates. *Saprolegnia parasitica* was used as outgroup. *A. astaci* APO3 sequence was retrieved from whole genome sequences made available on NCBI by the Broad Institute. Species, isolate ID, genotype and GenBank reference number are indicated. Lineages follow Diéguez-Uribeondo *et al.* (2009).

A. astaci isolate Pc (genotype D) consensus sequence was also compared to the one obtained from *A. astaci* isolate AP03 (genotype D) whole genome sequence made available on NCBI by the Broad Institute. The sequences were identical, suggesting that these intraspecies variants are consistent within genotype D.

3.2.6 Discussion

The analysis of ITS sequences and the phylogenetic tree (Figure 3.9) confirmed the identity of the working cultures and although it endorsed the efficacy of this marker to discriminate between morphologically similar species, the resolution below species level for *A. astaci* isolates was not possible (Figure 3.10) and the ITS marker was not able to distinguish between the genotypes defined by RAPD-PCR in this study and by Huang *et al.* (1994), Diéguez-Uribeondo *et al.* (1995) and Kozubíková *et al.* (2011).

An extensive review (Álvarez and Wendel, 2003) of ribosomal ITS sequences used in plant phylogenetic lists advantages and disadvantages of ITS to construct phylogenetic analysis. Some of the advantages are that the number of copies of this region in an individual is in the order of dozens to thousands and therefore it is easily isolated, and that, differently from the coding region of ribosomal DNA, ITS regions change more rapidly in the same individual due to insertions, deletions and point mutations and therefore they are a useful tool to discriminate species (Álvarez and Wendel, 2003). However, ITS regions are not always useful to distinguish genotypes within a single species, for example due to concerted evolution, a phenomenon that tends to homogenise differences in sequences raised in the same isolate or due to low genetic variation in pure laboratory cultures and in organisms that lack sexual reproduction (Álvarez and Wendel, 2003; Ganley and Kobayashi, 2007; Makkonen *et al.*, 2011).

The nucleotide variants noted in this study might be “true” variants or technical artefacts. True variants in different clones from pure cultures can be present because even if the concerted evolution tends to homogenise differences in sequences, it may not eliminate all variation immediately (Álvarez and Wendel, 2003; Simon and Weiß, 2008). Technical artefacts can be generated by the reagents and method used, for example, polymerase chain reactions are subjected to the occurrence of “PCR selection” (preferential amplification of a single sequence replicated during the first amplification cycles of the PCR), or to “PCR drift” (random events that occur in the first cycles of the reaction that can lead to erroneous outputs), events that can insert errors in the PCR product, which can lead to biased results (Álvarez and Wendel, 2003; Wagner *et al.*, 1994). In this study, the enzyme used for PCR was a Taq DNA polymerase,

which present an error rate of 8.0×10^{-6} errors per base due to the lack of proofreading activity (Cline *et al.*, 1996). These technical errors can be partly avoided by optimising the PCR reaction protocol, using a proofreading DNA polymerase, cloning the PCR products and screening multiple clones (Acinas *et al.*, 2005; Álvarez and Wendel, 2003).

Intragenomic and intraspecific nucleotide variants in ITS fragments in animal pathogenic *Aphanomyces* were noted also by Diéguez-Uribeondo *et al.* (2009) in sequenced PCR products compared to those retrieved from the NCBI Nucleotide database and by Makkonen *et al.* (2011) in sequenced PCR products and in cloned sequences. As extensively explained by Ganley and Kobayashi (2007), a true nucleotide variant should be found in the same variable site in more than one clone, while the intragenomic and intraspecific nucleotide variants in ITS sequences seen in this study are inconsistent. This suggests that either the variants observed in this study can be due to polymerase and PCR errors, or to a less efficient concerted evolution, already noted in ascomycetes and arbuscular mycorrhizal fungi (Pawlowska and Taylor, 2004; Simon and Weiß, 2008).

Like the ITS marker, the LSU phylogenetic tree confirmed its efficacy to discriminate between morphologically similar species but the resolution below species level for *A. astaci* isolates was not possible as the intragenomic and intraspecific nucleotide variants observed were inconsistent and due to polymerase or PCR errors and slow concerted evolution.

The main COI-based phylogenetic study for *Aphanomyces* spp. is that of Robideau *et al.* (2011), which focused mainly on plant pathogenic species. The only other available COI sequences are derived from the publicly available genome sequences generated by the Broad Institute (*i.e. A. astaci*). This study shows that COI can be a useful marker to discriminate between different *Aphanomyces* species and lineages (plant pathogens; animal pathogens; opportunistic/saprotrophic) and fits with the subdivision made by Diéguez-Uribeondo *et al.* (2009) based on the ITS marker. Moreover, this molecular marker can discriminate genotype D isolates from other *A. astaci* genotypes. Following Ganley and Kobayashi (2007), these nucleotide variants can be

considered “true variants” as appearing in all cloned sequences and from both *A. astaci* isolates belonging to genotype D.

These three studies combined show that the widely used molecular markers (ITS, LSU, and COI) are useful to distinguish closely related *Aphanomyces* species but they do not permit the distinction of *A. astaci* genotypes that could be instead distinguished using RAPD-PCR. The species identities of the cultures held in the laboratory have been confirmed by the three molecular markers, also confirming that *A. invadans*-like NJM9510 does not belong to the *A. invadans* species. The phylogenetic analysis shows that this isolate is instead closer related to the opportunistic/saprotrophic *Aphanomyces* species rather than to the animal parasitic ones and probably represents a new species. It was isolated in 1986 during an episode of epizootic ulcerative syndrome in Atlantic menhaden in the Pamlico River Estuary of North Carolina (USA) by Dykstra *et al.* (1986) and, in order to characterise the outbreak, it was arbitrarily chosen as a representative *Aphanomyces* between the ones isolated from fish lesions. In the Dykstra *et al.* (1986) study, the authors raised doubts about the correct ascription of an *Aphanomyces* isolate to an aquatic *Aphanomyces* species, as these organisms do not produce sexual stages under laboratory conditions and suggested to use molecular tools to clarify its identity, a suggestion followed up in this study.

As it was not possible to separate *A. astaci* genotypes with the three genetic markers used in this study, the identification of reliable tools to distinguish quickly and unambiguously *A. astaci* genotypes was still needed. Therefore, with the aid of WGS, a further investigation into the genetic differences between *A. astaci* genotypes was carried out and new genotyping tools developed (sections 3.3 and 3.4).

3.3 Use of genotype-specific genomic region as a tool to genotype *Aphanomyces astaci*

As previously mentioned, *Aphanomyces astaci* is a parasite of freshwater crayfish, involved in the decline of European indigenous crayfish species (Alderman *et al.*, 1984; Unestam, 1972). Five genotypes (A, B, C, D and E) have been detected and genetically characterised from pure cultures by RAPD-PCR, which is the preferred method to genotype this parasite (Diéguez-Uribeondo *et al.*, 1995; Huang *et al.*, 1994; Kozubíková *et al.*, 2011). Other fingerprinting tools developed to genotype this parasite include AFLP (Rezinciuc *et al.*, 2014)(Rezinciuc *et al.*, 2014)and microsatellites typing (Grandjean *et al.*, 2014; Rezinciuc *et al.*, 2014). Being able to quickly and reliably identify *A. astaci* genotypes is considerably important in epidemiological studies which proved to be a useful tool to better understand the history and spread of this disease in Europe (Lilley *et al.*, 1997a; Rezinciuc *et al.*, 2014; Viljamaa-Dirks *et al.*, 2013; Vrålstad *et al.*, 2014). The genotyping techniques currently available allow the distinction of all known *A. astaci* genotypes, but present some downsides that can influence the reproducibility and repeatability of the tests (section 1.3). As discussed in section 3.1, the need to apply the RAPD-PCR technique to pure cultures of *A. astaci* present some limitations, including the difficulty of interpretation of the band patterns and the standardisation of PCR reaction conditions (Jones *et al.*, 1997; Penner *et al.*, 1993; Skroch and Nienhuis, 1995). The investigation of an unusual RAPD-PCR profile obtained for *A. astaci* isolate 457 (section 3.1) enabled the identification of localised loss of heterozygosity in *loci* targeted by the RAPD-PCR that directly influenced the test (Figure 3.8). Moreover, variation in genotypes as defined by RAPD-PCR were already noted by other authors when characterising *A. astaci* with AFLP and microsatellites techniques (Grandjean *et al.*, 2014; Rezinciuc *et al.*, 2014; Viljamaa-Dirks *et al.*, 2014).

As it was not possible to separate *A. astaci* genotypes with the three genetic markers (internal transcribed spacer of the rDNA – ITS; D1-D2 regions of the large subunit ribosomal DNA – LSU; cytochrome oxidase subunit I - COI). previously tested in this study (section 3.2), the genetic diversity of *A. astaci* genotypes was further investigated, in order to develop PCR-based assays

centred on the presence/absence of a *genotype-specific* PCR product after amplification rather than the random amplification of DNA sequences.

3.3.1 Aims and objectives

This section describes the development of assays to genotype *A. astaci* by exploiting the WGS of 11 *A. astaci* isolates (Table 2.4). Bioinformatically, genotype-specific genomic regions were identified for each known genotype defined by RAPD-PCR (Diéguez-Uribeondo *et al.*, 1995; Huang *et al.*, 1994; Kozubíková *et al.*, 2011) and PCR assays were then developed to amplify these regions and distinguish the genotypes. These assays focus on the presence/absence of PCR product after amplification with genotype-specific primers. If the genotype of the isolate matches the genotype-specific primers, there is amplification of the genotype-specific region and consequently a PCR product. If the genotype does not match the primers, amplification is not possible and the PCR assay is negative. Moreover, to validate these assays, the primers were tested on pure cultures of *A. astaci* isolates belonging to the five known genotypes, other *Aphanomyces* species and Oomycetes (Table 2.1 and Table 2.3). Additionally, the assays were tested also on historical samples available in the Cefas laboratory (section 2.1.8) and on samples from a recent Italian crayfish plague outbreak (the reader is directed to section 3.5 for the results on the Italian outbreak).

3.3.2 Materials and methods

All materials and methods are described in detail in Chapter 2 – Materials and Methods. Briefly, gDNA was extracted from the 11 *A. astaci* isolates (Table 2.1) as described in section 2.3.2, submitted for WGS and sequenced using the ultra-high-throughput Illumina HiSeq 2500 System to generate paired 100, 200 or 150 bp reads (section 2.3.3). Prior to the *de novo* assembly, Illumina reads were quality checked and trimmed (section 2.4.3). To identify genotype-specific genomic regions, *de novo A. astaci* genomes were assembled with SPAdes (version 3.5.0) (Bankevich *et al.*, 2012), as described in section 2.4.4. A reference genome for each genotype was selected (section 2.4.10) on the basis

of completeness and contiguity, parameters assessed using BUSCO (Simão *et al.*, 2015) and QUAST (Gurevich *et al.*, 2013) (section 2.4.5). Paired-end reads from all *A. astaci* were aligned against the reference-genotype assembly using BWA (Li, 2013) (section 2.4.7). BEDTools (Quinlan and Hall, 2010) was then used to retrieve contigs with zero coverage (section 2.4.10). Contigs and sequences were visualised in IGV (Thorvaldsdóttir *et al.*, 2013) and genotype-specific primers designed (Table 2.2) and tested for specificity on gDNA extracted from all *Aphanomyces* isolates and from oomycetes (Table 2.1 and Table 2.3) following PCR protocols described in section 2.3.4.5. PCR fragments were checked by gel electrophoresis (section 2.3.1), positive PCR products purified (section 2.3.5) and sent to Eurofins MWG Operon commercial sequencing facility to sequence (2.3.8). Sequences were visualised and aligned with BioEdit version 7.0.8 (Hall, 1999) to confirm the sequences of the genotype-specific genomic regions identified in the assembled genomes. Primers sensitivity was validated on Cefas crayfish historical samples (section 2.1.8).

3.3.3 The genome-informed genotyping markers successfully discriminate *A. astaci* genotypes

For each *A. astaci* reference-genotype assembly, genotype-specific genomic regions were identified, the corresponding contigs visualised in IGV and primers designed. Figure 3.14 and Figure 3.15, extrapolated from IGV, show the alignments of the 12 *A. astaci* paired-end reads to the reference genomes. The grey bar chart represents the coverage track, which displays the depth of reads coverage for each genomic position for each isolate against a reference genome. The regions selected to develop the primers had no (Figure 3.14 and Figure 3.15, genotypes A, B, C and E) or low level (Figure 3.14, genotype D) of heterozygosity within the paired-ends reads and no (Figure 3.14 and Figure 3.15, genotypes B, C, D and E) or limited (Figure 3.15, genotype A) paired-ends reads coverage for the isolates belonging to a different genotype against the reference genome. Genotype-specific genomic regions of ~800 bp were selected to facilitate visualisation of PCR products on agarose gels after amplification. Genotype-specific primer pairs for each genotype were designed

(Table 2.2) and tested on *Aphanomyces* and other oomycetes gDNA extractions (Table 2.1 and Table 2.3).

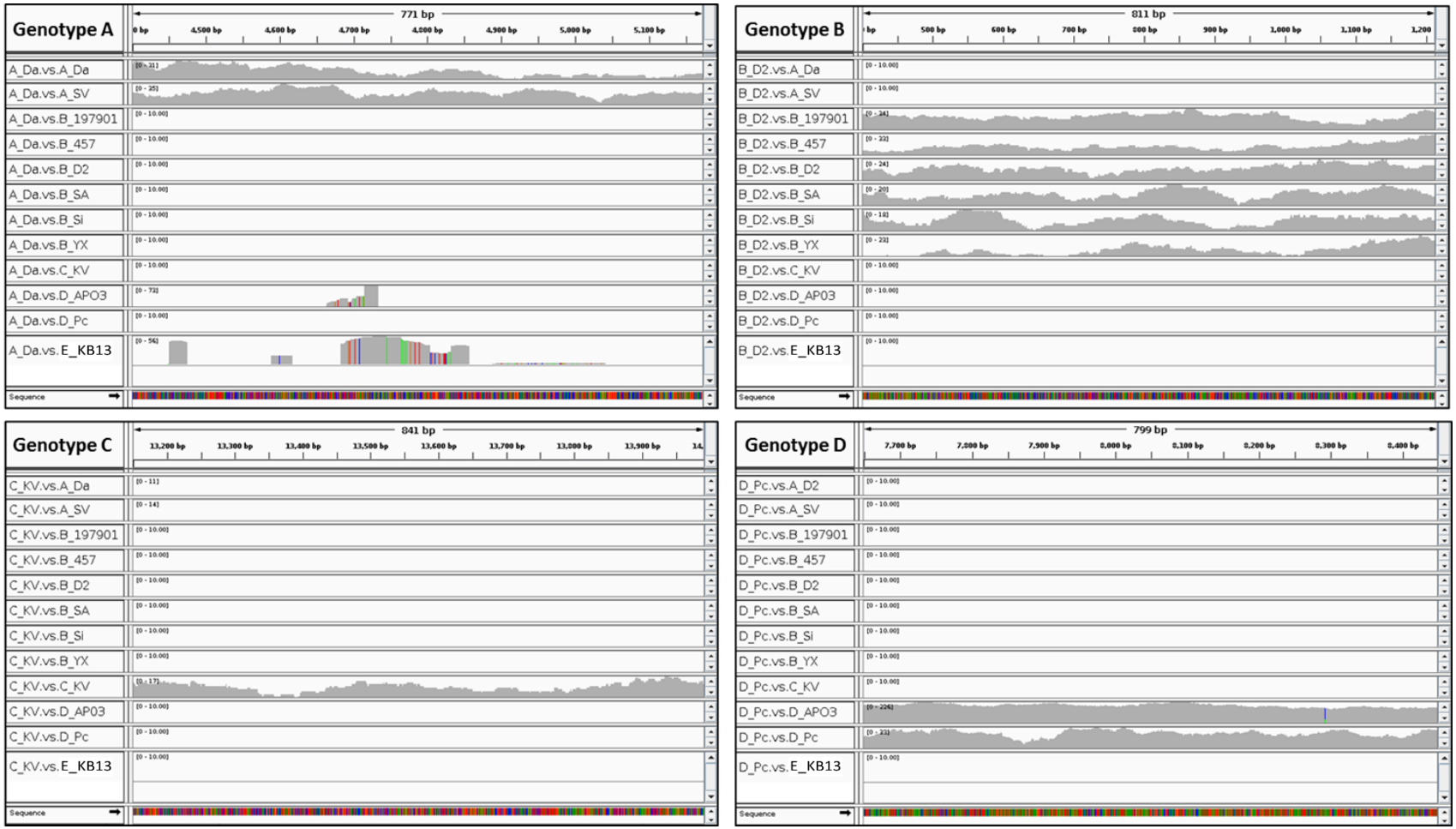


Figure 3.14 *A. astaci* genotype-specific genomic regions visualised in IGV for genotypes A, B, C, and D. Reference genomes: genotype A: *A. astaci* Da; genotype B: *A. astaci* D2; genotype C: *A. astaci* KV; genotype D: *A. astaci* Pc.. From top to bottom, aligned paired-end reads of: *A. astaci* Da (genotype A); SV (genotype A); 197901 (genotype B); 457 (genotype B); D2 (genotype B); SA (genotype B); Si (genotype B); YX (genotype B); KV (genotype C); APO3 (genotype D); Pc (genotype D); KB13 (genotype E). Primers were designed at the extremities of shown regions. Individual vertical bars: depth of read coverage of each genomic position against the reference genome. Grey bars: majority (> 80 %) of nucleobases aligned match the reference genome. Green bars: at least 20 % of nucleobases aligned not matching the reference genome and substituted by adenosine (A).

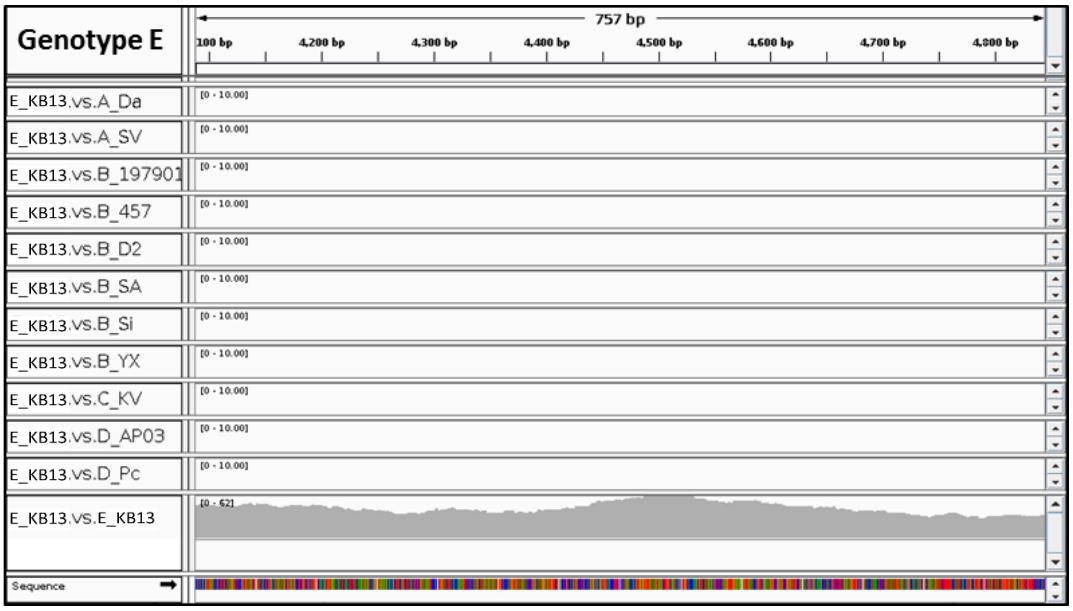


Figure 3.15 *A. astaci* genotype-specific genomic regions visualised in IGV for genotype E. Reference genomes for genotype E: *A. astaci* KB13. From top to bottom, aligned paired-end reads of: *A. astaci* Da (genotype A); SV (genotype A); 197901 (genotype B); 457 (genotype B); D2 (genotype B); SA (genotype B); Si (genotype B); YX (genotype B); KV (genotype C); APO3 (genotype D); Pc (genotype D); KB13 (genotype E). Primers were designed at the extremities of shown regions. Individual vertical bars: depth of read coverage of each genomic position against the reference genome. Grey bars: majority (> 80 %) of nucleobases aligned match the reference genome. Green bars: at least 20 % of nucleobases aligned not matching the reference genome and substituted by adenosine (A).

For genotype A, the fragment amplified by the genotype-specific primer pair "A_unique" was 765 bp long (Figure 5.3) (including primer sequences) and the amplification was produced only in *A. astaci* isolates belonging to genotype A (Figure 3.16 A). No amplification was noted from *A. astaci* belonging to other genotypes and in other *Aphanomyces* species, nor in other oomycete isolates (Figure 3.16 A and B).

As for the previous test, the PCR with genotype-specific primer pair for genotype B "B_unique" produced positive results only in *A. astaci* isolates belonging to genotype B (Figure 3.16 C) and no amplification was present in other *Aphanomyces* and oomycete isolates (Figure 3.16 C and D). The fragment amplified by the genotype-specific primers was 884 bp long, including primer sequences (Figure 5.4).

Amplification of the gDNA with genotype-specific primer pair for genotype C "C_unique" produced positive results in *A. astaci* isolate belonging to genotype C (Figure 3.16 E). The fragment length was 771 bp, including primer sequences (Figure 5.5). No amplification was present in other *Aphanomyces* isolates

(Figure 3.16 E), but bands were detected for *P. flevoense* and *Phoma*-like gDNA extraction (Figure 3.16 F). As the PCR products for *Phoma*-like isolate were of different sizes in respect to the genotype C control PCR product, the investigation was not carried forward. For *P. flevoense*, the PCR product was purified, cloned, sequenced and aligned to the reference genotype C sequence. From the alignment, *P. flevoense* and *A. astaci* genotype C share the same forward primer site, while the rest of sequence composition is different and it was not possible to match the reverse primer (Figure 5.5).

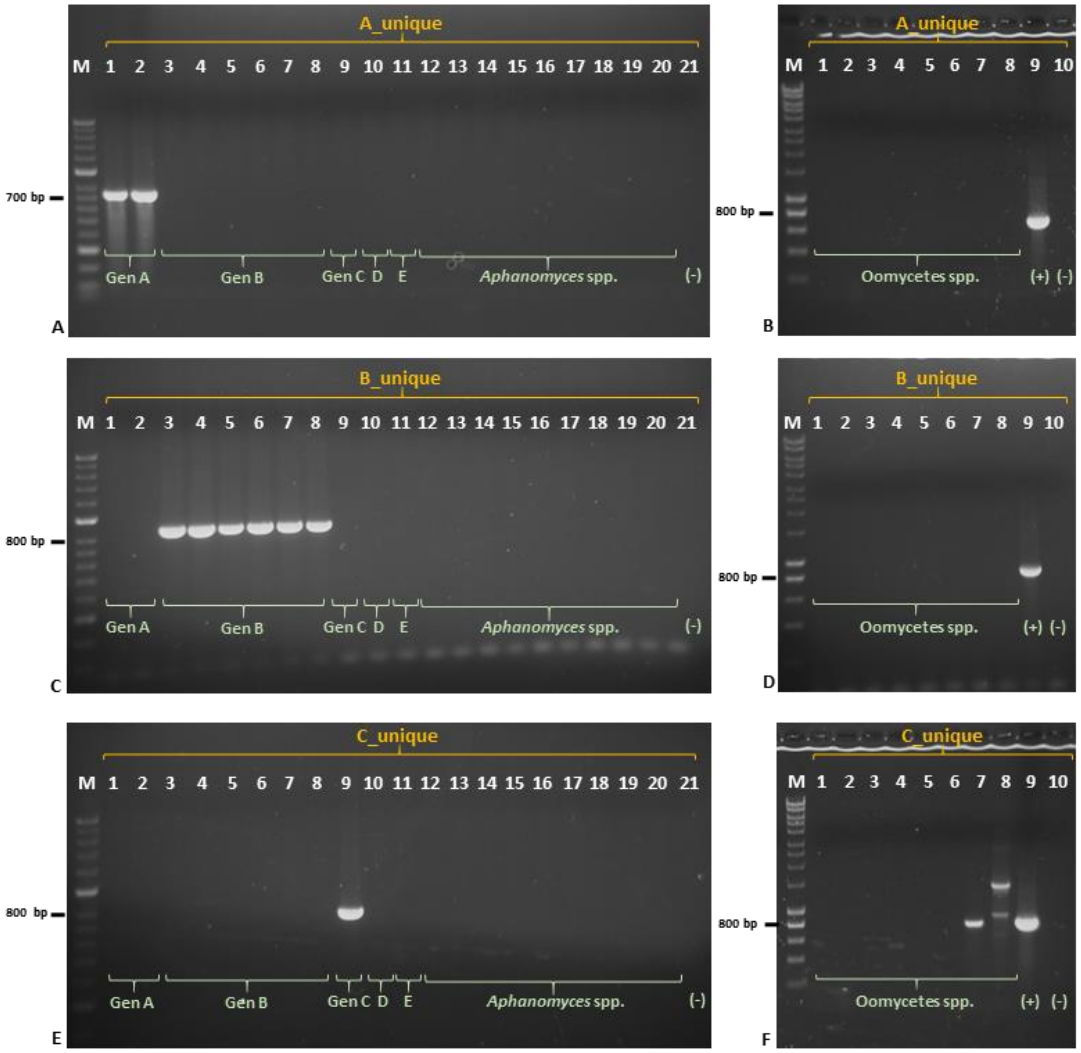


Figure 3.16 Specificity test for genotype A, B, and C specific primer pairs on *Aphanomyces* and other oomycetes isolates. **A**, **C** and **E**: M, Bioline HyperLadder II; 1, *A. astaci* SV (genotype A); 2, *A. astaci* Da (genotype A); 3, *A. astaci* 457 (genotype B); 4, *A. astaci* 197901 (genotype B); 5, *A. astaci* SA (genotype B); 6, *A. astaci* D2 (genotype B); 7, *A. astaci* YX (genotype B); 8, *A. astaci* Si (genotype B); 9, *A. astaci* KV (genotype C); 10, *A. astaci* Pc (genotype D); 11, *A. astaci* KB13 (genotype E); 12, *A. invadans* NJM9701; 13, *A. invadans* NJM8997; 14, *A. invadans* NJM9030; 15, *A. invadans* NJM0002; 16, *A. invadans* GWR; 17, *A. invadans*-like NJM9510; 18, *A. frigidophilus*, AP5; 19, *A. frigidophilus* RP1; 20, *A. frigidophilus*d RP2; 21, negative control. **B**, **D** and **F**: M, Bioline HyperLadder I; 1, *S. parasitica*; 2, *S. furcata*; 3, *L. caudata*; 4, *A. racemosa*; 5, *A. laevis*; 6, *P. monospermum*; 7, *P. flevoense*; 8, *Phoma*-like; 9**B**, *A. astaci* Da (genotype A); 9**D**, *A. astaci* D2 (genotype B); 9**F**, *A. astaci* KV (genotype C); 10, negative control. (+), positive control; (-) negative control.

Similarly to A and B genotype-specific assays, the genotype-specific primers for genotype D “D_unique” and E “E_unique” amplified the selected genomic region in *A. astaci* isolate belonging to genotype D (Figure 3.17 A) and E respectively (Figure 3.17 D), without amplification in other *Aphanomyces* (Figure 3.17 A, B, D and E) or oomycetes isolates (Figure 3.17 C and F). The fragment amplified by the genotype-specific primers were 761 bp long for genotype D (Figure 5.6) and 736 bp long for genotype E (Figure 5.7) including primer sequences.

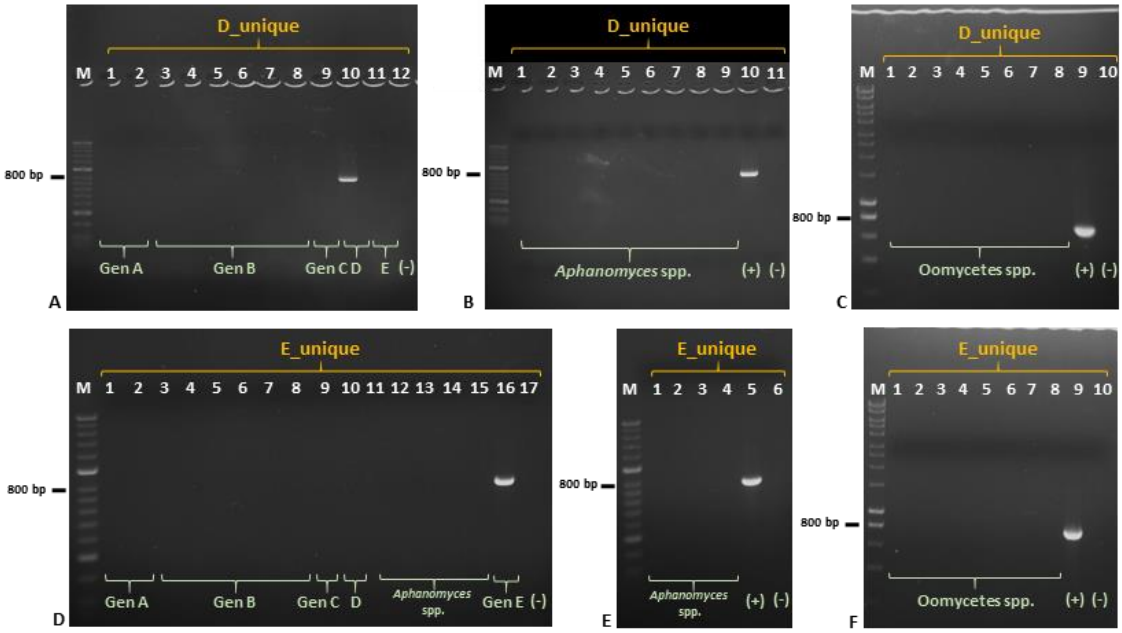


Figure 3.17 Specificity test for genotype D and E specific primer pairs on *Aphanomyces* and other oomycetes isolates. **A**: M, Bioline HyperLadder II; 1, *A. astaci* SV (genotype A); 2, *A. astaci* Da (genotype A); 3, *A. astaci* 457 (genotype B); 4, *A. astaci* 197901 (genotype B); 5, *A. astaci* SA (genotype B); 6, *A. astaci* D2 (genotype B); 7, *A. astaci* YX (genotype B); 8, *A. astaci* Si (genotype B); 9, *A. astaci* KV (genotype C); 10, *A. astaci* Pc (genotype D); 11, *A. astaci* KB13 (E); 12, negative control. **B**: M, Bioline HyperLadder II; 1, *A. invadans* NJM9701; 2, *A. invadans* NJM8997; 3, *A. invadans* NJM9030; 4, *A. invadans* NJM0002; 5, *A. invadans* GWR; 6, *A. invadans*-like NJM9510; 7, *A. frigidophilus* AP5; 8, *A. frigidophilus* RP1; 9, *A. frigidophilus* RP2; 10, *A. astaci* Pc; 11, negative control. **C** and **F**: M, Bioline HyperLadder I; 1, *S. parasitica*; 2, *S. furcata*; 3, *L. caudata*; 4, *A. racemosa*; 5, *A. laevis*; 6, *P. monospermum*; 7, *P. flevoense*; 8, *Phoma*-like; 9**C**, *A. astaci* Pc; 9**F**, *A. astaci* KB13; 10, negative control. **D**: M, Bioline HyperLadder II; 1, *A. astaci* SV (genotype A); 2, *A. astaci* Da (genotype A); 3, *A. astaci* 457 (genotype B); 4, *A. astaci* 197901 (genotype B); 5, *A. astaci* SA (genotype B); 6, *A. astaci* D2 (genotype B); 7, *A. astaci* YX (genotype B); 8, *A. astaci* KV (genotype C); 9, *A. astaci* Pc (genotype D); 10, *A. invadans* NJM9701; 11, *A. invadans* NJM9030; 12, *A. invadans* NJM0002; 13, *A. invadans* GWR; 14, *A. frigidophilus* RP2; 15, *A. frigidophilus* AP5; 16, *A. astaci* KB13; 17, negative control. **E**: M, Bioline HyperLadder II; 1, *A. astaci* Si (genotype B); 2, *A. invadans* NJM8997; 3, *A. invadans*-like NJM9510; 4, *A. frigidophilus* RP1; 5, *A. astaci* KB13; 6, negative control. (+), positive control; (-) negative control.

The PCR assays were tested on Cefas outbreak samples (section 2.1.8) in which *A. astaci* was detected by amplification and sequencing of ITS regions. These samples have a low concentration of *A. astaci* gDNA in comparison to the gDNA extracted from pure cultures and, unfortunately, the sensitivity of the

assay was too low to detect *A. astaci*, resulting in negative PCR. Increasing the PCR cycles number from 35 to 55 or applying a touchdown PCR (section 2.3.4.7) was still unsuccessful (results not shown).

To increase the sensitivity of the PCR-based assays, a similar bioinformatics approach was used to align *A. astaci* paired-end reads from the 11 sequenced isolates against an *A. astaci* mitochondrial DNA assembly as the number of copies of the mitochondrial genome is higher in eukaryotic cells in comparison to the nuclear genome. On the other hand, the mitochondrial genome is smaller in comparison to the nuclear genome and, as expected, it was not possible to find genotype-specific regions amplified only by a set of primers for each known genotype.

3.3.4 Discussion

This study demonstrates that bioinformatics methods and information on genomes are valuable and useful tools to develop informative molecular markers and that combining this information with molecular biology techniques is possible to genotype the causative agent of crayfish plague, *A. astaci*. By further investigating the genetic diversity of *A. astaci* genotypes (as defined by RAPD-PCR), genotype-specific genomic regions were identified and genotyping PCR based assays were developed. Homozygous sites were selected over heterozygous ones to limit primers misaligning that could influence the PCR amplification, resulting in presence of aspecific bands or complete absence of PCR products and primers matching the homozygous sites of the genomic region were selected.

The RAPD-PCR approach does not require the knowledge of the DNA sequences that are going to be amplified (Harris, 1999). Instead, the importance of knowing the targeted sequence when genotyping has risen after detection of *A. astaci* 457 unusual genotype B RAPD-PCR profile due to localised loss of heterozygosity in *loci* amplified by the RAPD-PCR and characteristic of the other isolates belonging to genotype B. These newly developed PCR based assays focus on the presence/absence of PCR product after amplification with genotype-specific primers. If the genotype of the isolate matches the genotype-

specific primers, there is amplification of the genotype-specific region and consequently a PCR product. If the genotype does not match the primers, amplification is not possible and the PCR assay is negative. The assays were tested on *Aphanomyces* and on other oomycete isolates (Table 2.1 and Table 2.3). The cross reaction of the primers with other genomic regions outside the targeted ones was limited. The assays could easily distinguish *A. astaci* genotypes, not only bioinformatically, but also from pure culture genomic extractions. The only cross-reaction noted was the one between the genotype C specific primer pair with isolates *Phoma*-like and *P. flevoense* (Figure 3.16 F). As the bands of *Phoma*-like were of different sizes in respect of the positive control, the investigation was not carried forward as in a hypothetical laboratory screening situation, bands of different sites not matching the control would be discarded. Instead, *P. flevoense* and *A. astaci* genotype C share the same forward primer site, while the rest of sequence composition is different and it was not possible to match the reverse primer (Figure 5.5). This artefact might be prevented by optimisation of PCR conditions (*i.e.* increasing primer aligning temperature). This finding confirms again that the genotype C specific region is still specific and useful to distinguish *A. astaci* genotype C.

Unfortunately, the primers did not work on the Cefas outbreak samples (2.1.8), positive for *A. astaci* by amplification and sequencing of ITS regions. This might be due to the low copy number of the selected genomic region in contrast to the ITS regions, repeats that are present numerously in the nuclear DNA (Álvarez and Wendel, 2003). These samples have a low concentration of *A. astaci* gDNA in comparison to the gDNA extracted from pure cultures and the sensitivity of the assay does not permit to detect such low concentration of *A. astaci* gDNA, resulting in negative PCRs. From the analysis of the whole genome assemblies of *A. astaci*, the ITS regions in this species are about 60 times more represented in the nuclear DNA in comparison to single-copy genes (section 3.6.5). Another most likely explanation for the lack of amplification from the Cefas outbreak samples is the possible degradation of the DNA in those samples due to storing conditions or enzymatic degradation (Wilson, 1997). The PCR product that would be generated using the genotype specific primers is significantly long, making the genotype specific primers less likely to generate a product when using samples that have been stored for a long time. Even if the

sensitivity of the assay is low and positives results were present only on pure cultures, the markers here developed are promising and could substitute or integrate the RAPD-PCR technique on pure cultures. To reinforce the efficacy of the assays, more *A. astaci* isolates and population should be tested, especially for isolates belonging to genotypes C and E, the most closely related to genotype B from the analysis of SNVs (section 3.1).

The development of these markers has some advantage in respect to the RAPD-PCR assay. For example, the reproducibility between laboratories should be higher as the primers designed are specific for each known genotype and target a specific region of the DNA instead of randomly amplify fragments. Moreover, the PCR product can be sequenced, adding quality to the data, and directly compared to the genotype-specific region sequences to easily detect homologies (or differences) between the sequences.

3.4 Use of genotype-specific genomic single nucleotide variants as a tool to genotype *Aphanomyces astaci* relying on enzymatic assay

In epidemiology, being able to identify *A. astaci* genotypes directly from host tissues is an important step to understand the spread of crayfish plague in Europe, to prevent new epizootic events and to track possible source of infections.

While RAPD-PCR, AFLP and microsatellites allow the distinction of all known *A. astaci* genotypes, they present some downsides that can influence the reproducibility and repeatability of the tests (sections 1.3, 3.1 and 3.3), therefore the identification of reliable markers is still needed. Ideally, these markers should not rely on culture methods, differentiate between species, and detect different genotypes and strains that can be present at the same time in the same sample. Section 3.2 shows that widely used molecular markers (internal transcribed spacer of the rDNA – ITS; D1-D2 regions of the large subunit ribosomal DNA – LSU; cytochrome oxidase subunit I - COI) are useful to distinguish among closely related *Aphanomyces* species but do not resolve *A. astaci* intraspecies genotypes. Only the COI molecular marker displayed a higher resolution level, successfully distinguishing *A. astaci* genotype D from the other genotypes, but not sufficient to discriminate the other four genotypes. In section 3.3 the genetic diversity of *A. astaci* genotypes was further investigated, leading to the development of genetic markers used for diagnostic and epidemiological studies. Unfortunately, the primers did not work on Cefas outbreak samples possibly due to the low sensitivity of the assay which does not permit to detect low concentration of *A. astaci* gDNA.

3.4.1 Aims and objectives

In the work presented below, *A. astaci* assemblies and WGS obtained in this study were further exploited to develop more sensitive genotyping techniques, by identifying genotype-specific SNVs in both nuclear and mtDNA that can unequivocally distinguish *A. astaci* genotypes. Primers were designed to amplify the regions containing the SNVs and a restriction digestion assay developed to

distinguish the genotypes. To validate these assays, the primers were tested on *A. astaci* isolates belonging to the five known genotypes, other Oomycetes (Table 2.1 and Table 2.3) and infected crayfish and historical samples available at the Cefas laboratory (Table 5.1 and Table 5.2). Moreover, as for the genotype-specific genomic region assay, the SNVs assays were tested on samples from a recent Italian crayfish plague outbreak (results in section 3.5).

3.4.2 Materials and methods

All methods and materials are described in detail in Chapter 2 – Materials and Methods. Briefly, gDNA was extracted from the 11 *A. astaci* isolates held in Cefas OCC (Table 2.1) (section 2.3.2), submitted for WGS and sequenced using the ultra-high-throughput Illumina HiSeq 2500 System to generate paired 100 bp or 150 bp reads (section 2.3.3). Illumina reads were quality checked and trimmed (section 2.4.3). The generate a list of *A. astaci* genotype-specific SNVs, trimmed pair-end reads from each sequenced *A. astaci* isolate were aligned against a reference genome with BWA (Li, 2013) and SAMtools (Li *et al.*, 2009) (section 2.4.7). Reference genome were: *A. astaci* APO3 assembly (GenBank assembly accession number: GCA_000520075.1) to detect SNVs in the nuclear DNA; *A. astaci* mitochondrial genome (*de novo* assembly as described in 2.4.6) to detect SNVs in the mtDNA. SNVs were called as per section 2.4.8. Every genomic position of the reference genomes was bioinformatically defined as ambiguous *(e.g.* due to heterozygosity, or insufficient data) or unambiguous (*e.g.* sufficient depth and consensus of aligned sequence data. Genotype-specific SNVs in restriction sites were detected as per section 2.4.9. Flanking sequences and SNVs were visualised in IGV and primers to amplify the fragments designed (section 2.1.2). Specificity of the primers was tested on gDNA extracted from all *Aphanomyces* isolates and oomycetes (Table 2.1 and Table 2.3) following PCR protocols described in sections 2.3.4.6 and 2.3.4.8. PCR fragments were checked by gel electrophoresis (section 2.3.1), positive PCR products purified (section 2.3.5) and sent to Eurofins MWG Operon commercial sequencing facility to sequence (2.3.8). Sequences were visualised and aligned with BioEdit version 7.0.8 (Hall, 1999) to confirm the presence (or the absence) of the SNVs in the restriction site. PCR fragments obtained with the genotype-specific nuclear and

mitochondrial SNVs primers were digested following the protocols described in section 2.3.9. Digested products were checked by gel electrophoresis using 2 % agarose gel stained with Midori Green Advance stain. A semi-nested PCR assay (section 2.3.4.9) was developed in order to exploit the SNVs in the region amplified by B_MITO primers as it was not possible to find a restriction enzyme able to digest the PCR product at the level of the SNV. Primers sensitivity was validated on Cefas crayfish historical samples (section 2.1.8).

3.4.3 Development of restriction digestion assays to detect and discriminate all known *Aphanomyces astaci* genotypes targeting nuclear DNA

The output of this part of the study is a restriction digestion assay to detect and distinguish *A. astaci* genotypes; it is based on the amplification by PCR of a targeted nuclear DNA region containing the selected SNV in the restriction site. By digesting the PCR product, the isolate's genotype can be determined by the pattern of bands on the agarose gel. The development of these markers was enabled by WGS of representatives of each genotype A, B, C, D, and E, and by generating a list of *A. astaci* genotype-specific SNVs in restriction sites.

To identify potential SNV markers for this purpose, each genomic position of the *A. astaci* APO3 (genotype D) genome assembly was categorised as ambiguous (*e.g.* due to heterozygosity, or insufficient or contradictory data) or unambiguous (*e.g.* sufficient depth and high degree of consensus in the aligned sequence data) and for each *A. astaci* genotype, genotype-specific SNVs in restriction sites were identified *in silico*, visualised in IGV (Figure 3.18: genotype A; Figure 3.19: genotypes B, C, D, and E) and primers to amplify the fragments were designed. Table 3.1 (section 3.1) lists the numbers of genotype-specific SNVs found in the whole genome of each genotype. Both figures below show the alignments of the 12 sets of *A. astaci* paired-end reads to *A. astaci* APO3 reference assembly. Each individual vertical bar represents the coverage track, which displays the depth of read coverage for each genomic position against a reference genome. A grey bar indicates that the majority of nucleobases aligned in that position to the reference genome is matching the reference genome while a bar of different colour indicates that the nucleobases aligned in that

position do not match the reference genome. Green refers to the nucleobase variant adenosine (A), red to tyrosine (T), yellow to guanosine (G) and blue to cytosine (C). For example, in Figure 3.18 the first two alignments represent *A. astaci* Da and SV (both genotype A) paired-end reads aligned to the reference genome. In the yellow box, the reference genome sequence (on the bottom of the figure) is “GCAC”, same as all the other isolates aligned (grey bars), while *A. astaci* Da and SV sequence is “GCGC” as the nucleobase “A” is substituted with the variant “G”, sequence cut by the restriction enzyme HhaI. In Figure 3.19, isolates belonging to genotype B are characterised by the presence of the nucleobase “C” instead of “A”, genotype C by nucleobase “G” instead of “A”, genotype D by nucleobase “G” instead of “T” and genotype E by nucleobase “C” instead of “T”. For all genotypes, the detected genotype-specific SNVs chosen were in the restriction site cut by HhaI. Bars presenting two different colours (*e.g.* Figure 3.19, genotypes D and E) indicates the occurrence of reads with both (or more) nucleobases for the same genomic site, caused for example by the presence of heterozygosity. These positions, categorised as ambiguous, are discarded by the script used to detect genotype-specific SNVs in restriction sites.

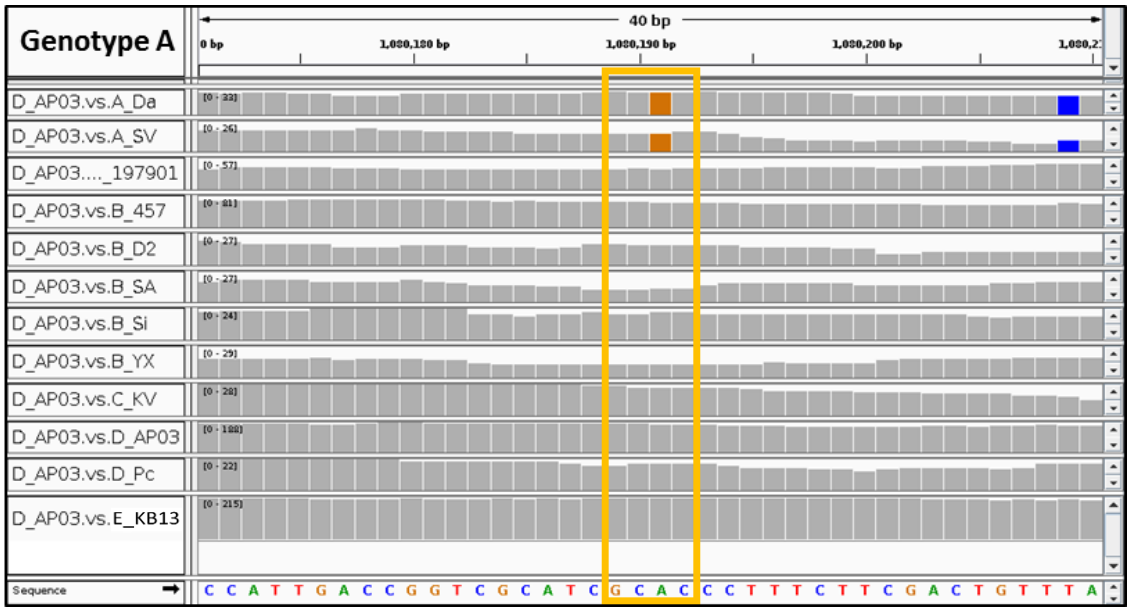


Figure 3.18 A nuclear SNV unique to *A. astaci* genotype A that falls within a restriction site. An alignment of genomic sequence reads aligned against a reference genome is visualised in IGV. Yellow box: SNVs in restriction site cut by HhaI (recognition sequence: GCGC). From top to bottom, aligned paired-end reads of: *A. astaci* Da (genotype A); SV (genotype A); 197901 (genotype B); 457 (genotype B); D2 (genotype B); SA (genotype B); Si (genotype B); YX (genotype B); KV (genotype C); APO3 (genotype D); Pc (genotype D); KB13 (genotype E). Reference genome "Sequence": *A. astaci* APO3 (genotype D). Individual vertical bars: depth of read coverage of each genomic position against the reference genome. Grey bars: majority (> 80 %) of nucleobases aligned match the reference genome. Yellow and blue bars: at least 20 % of nucleobases aligned not matching the reference genome and substituted respectively by guanosine (G) and cytosine (C).

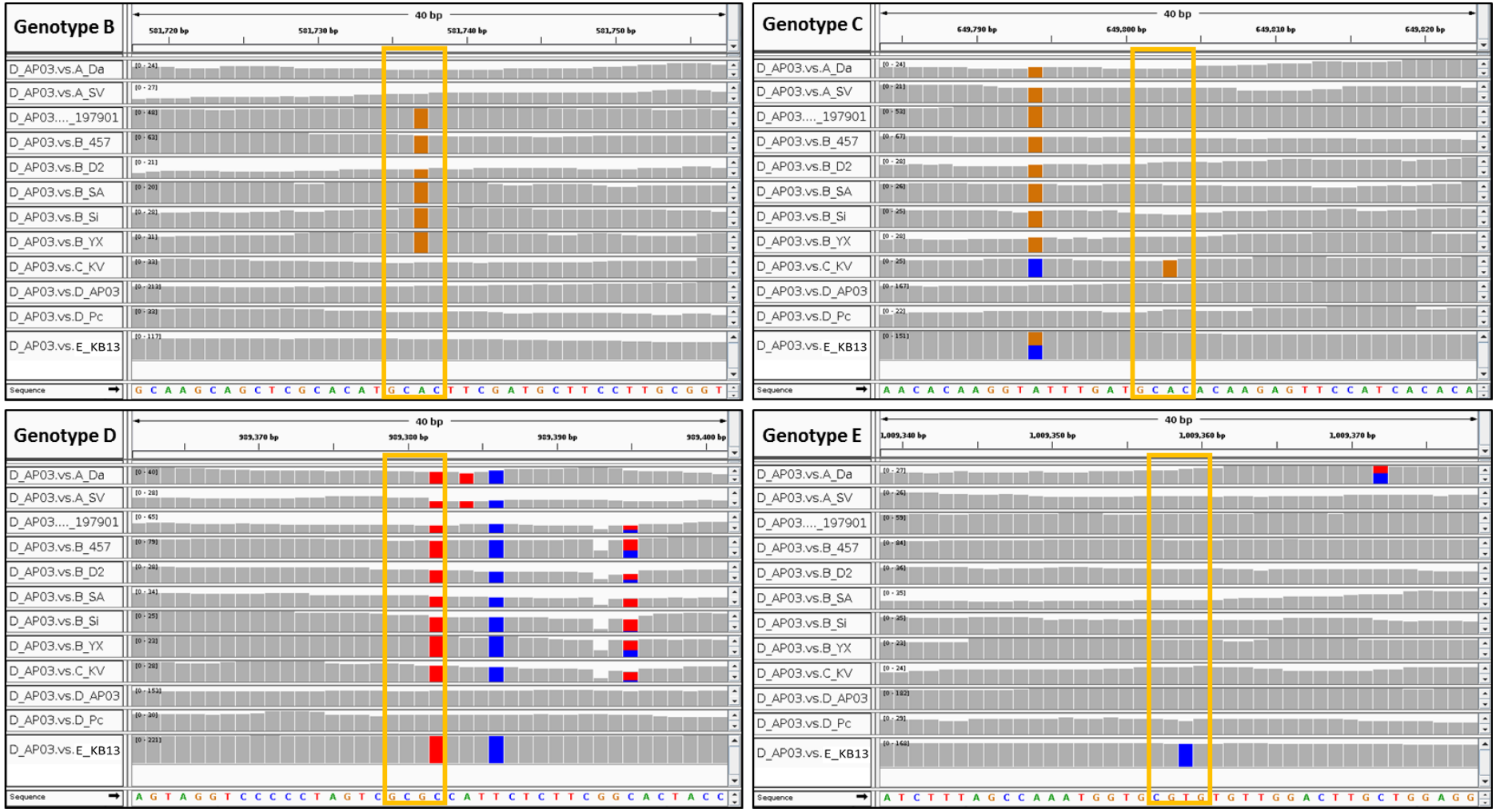


Figure 3.19 Nuclear SNVs uniques to *A. astaci* genotype B, C, D, and E that fall in restriction site. Alignments of genomic sequence reads against a reference genome visualised in IGV. Yellow boxes: SNVs in restriction site cut by HhaI (recognition site: HhaI). From top to bottom, aligned paired-end reads of: *A. astaci* Da (genotype A); SV (genotype A); 197901 (genotype B); 457 (genotype B); D2 (genotype B); SA (genotype B); Si (genotype B); YX (genotype B); KV (genotype C); APO3 (genotype D); Pc (genotype D); KB13 (genotype E). Reference genome "Sequence": *A. astaci* APO3 (genotype D). Individual vertical bars: depth of read coverage of each genomic position against the reference genome. Grey bars: majority (> 80 %) of nucleobases aligned match the reference genome. Red, yellow, and blue bars: at least 20 % of nucleobases aligned not matching the reference genome and substituted respectively by tyrosine (T), guanosine (G) and cytosine (C).

To confirm whether the restriction-site polymorphisms predicted with the bioinformatics approach would be reflected in the *in vitro* restriction digestion assays, primers to amplify the regions containing the SNVs were designed at ~200 bp before and after the SNVs (final fragment of ~400 bp) to facilitate visualisation of PCR products on agarose gels after amplification and after digestion and primer pairs A_HhaI and B_HhaI were firstly used on some of the *A. astaci* isolates. In particular, the primer pair A_HhaI was used to amplify the genomic region containing the SNV from *A. astaci* SV and Da (genotype A) gDNA, while primer pair B_HhaI was used on *A. astaci* 457, 197901, SA, D2, YX and Si (genotype B). The PCR products (Figure 3.20 A) were then subjected to enzymatic digestion with *Hha*I (Figure 3.20 B). As expected, after enzymatic digestion, the digested PCR products were half of the size of the original PCR products, thus the *in vitro* assays reflected the *in silico* assays. In Figure 3.20 B, partial digestions of the PCR products (faint bands) can be seen in some of the lanes, especially lane number 7. Increasing the duration of the digestion from one to two hours was enough to have a complete digestion of the PCR product.

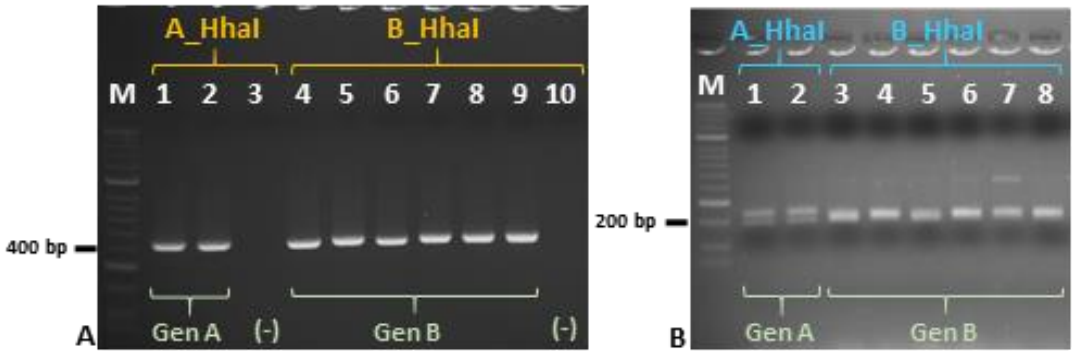


Figure 3.20 *A. astaci* PCR products amplified with A_HhaI and B_HhaI primer pairs (**A**) and their restriction digestion (**B**). **A**: 1, *A. astaci* SV (genotype A); 2, *A. astaci* Da (genotype A); 3, negative control; 4, *A. astaci* 457 (genotype B); 5, *A. astaci* 197901 (genotype B); 6, *A. astaci* SA (genotype B); 7, *A. astaci* D2 (genotype B); 8, *A. astaci* YX (genotype B); 9, *A. astaci* Si (genotype B); 10, negative control. **B**: 1, *A. astaci* SV (genotype A); 2, *A. astaci* Da (genotype A); 3, *A. astaci* 457 (genotype B); 4, *A. astaci* 197901 (genotype B); 5, *A. astaci* SA (genotype B); 6, *A. astaci* D2 (genotype B); 7, *A. astaci* YX (genotype B); 8, *A. astaci* Si (genotype B). M, Bioline HyperLadder II. (-), negative control.

The primer pairs and the restriction digestion assays for all genotypes were then tested on the remaining *A. astaci* isolates and the other *Aphanomyces* and oomycetes isolates available in Cefas OCC (Table 2.1 and Table 2.3) and PCR products digested.

The primer pair A_HhaI amplified DNA regions in all *A. astaci* isolates (Figure 5.8 A) but no amplification was noted for other *Aphanomyces* species (Figure 5.8 B) and for the oomycetes tested (Figure 5.8 C). Figure 3.21 shows the results of the restriction digestion assay on the PCR products from *A. astaci* isolates. In this figure, PCR products subjected to enzymatic digestion are indicated with the blue arrows, while the original (not subjected to enzymatic digestion) PCR products are indicated by the yellow arrows. The enzymatic digestion occurred only for the PCR product of isolate number 10, which is *A. astaci* Da (genotype A), while it did not occur for the isolates belonging to the other genotypes.

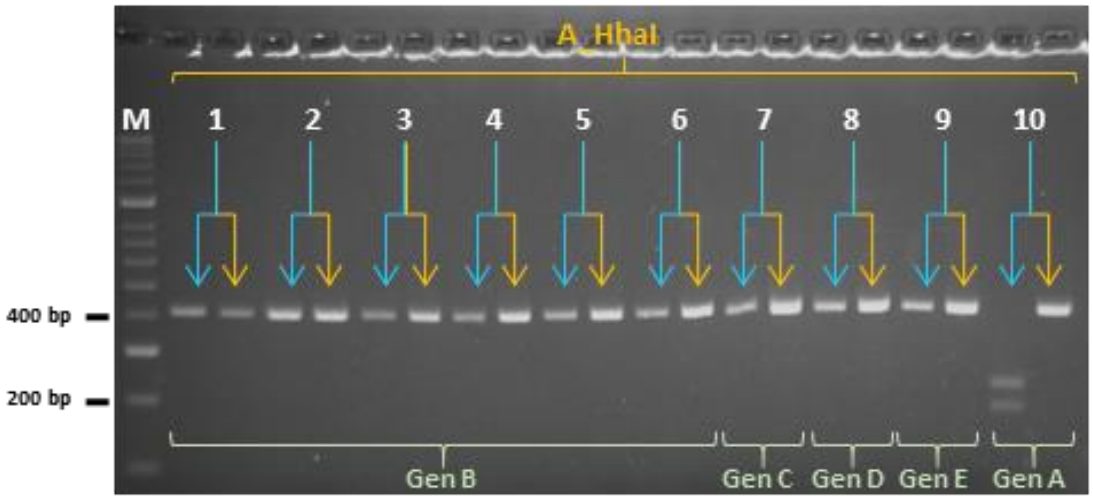


Figure 3.21 Restriction digestions of PCR products from the remaining *A. astaci* isoltes amplified with primer pair A_HhaI. M, Bioline HyperLadder II; 1, *A. astaci* 457 (genotype B); 2, *A. astaci* 197901 (genotype B); 3, *A. astaci* SA (genotype B); 4, *A. astaci* YX (genotype B); 5, *A. astaci* D2 (genotype B); 6, *A. astaci* Si (genotype B); 7, *A. astaci* KV (genotype C); 8, *A. astaci* Pc (genotype D); 9, *A. astaci* KB13 (genotype E); 10, *A. astaci* Da (genotype A). Blue arrows: PCR product after enzymatic digestion. Yellow arrows: original PCR product.

The primer pair B_HhaI amplified DNA regions for all *A. astaci* isolates (Figure 5.9 A) tested, no amplification was noted for other *Aphanomyces* species (Figure 5.9 C). However, the primers amplified genomic regions from some oomycetes (Figure 5.9 D). As the bands resulting from the PCR were of different sizes to the one in *A. astaci* D2 (genotype B) lane 9 (Figure 5.9 D), the restriction digestion assay was not applied as the products can be distinguished already by size and discarded as not relevant. Figure 3.22 shows the results of the restriction digestion on the PCR products amplified from *A. astaci* isolates. The enzymatic digestion occurred only for the PCR product in well number 6, which is *A. astaci* D2 (genotype B), while it did not occurr for isolates belonging to other genotypes.

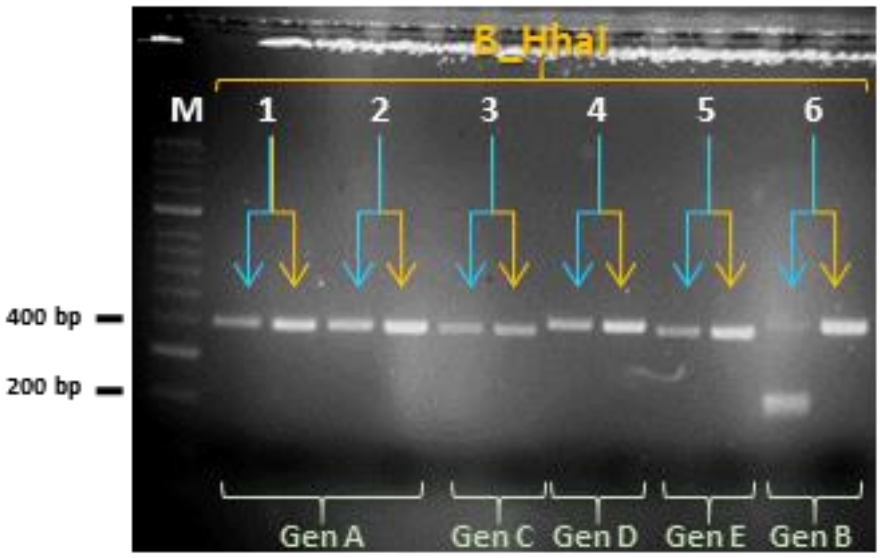


Figure 3.22 Restriction digestions of PCR products from the remaining *A. astaci* isolates amplified with primer pair B_HhaI. M, Bioline HyperLadder II; 1, *A. astaci* SV (genotype A); 2, *A. astaci* Da (genotype A); 3, *A. astaci* KV (genotype C); 4, *A. astaci* Pc (genotype D); 5, *A. astaci* KB13 (genotype E); 6, *A. astaci* D2 (genotype B). Blue arrows: PCR product after enzymatic digestion. Yellow arrows: original PCR product.

The primer pair C_HhaI amplified DNA regions of all *A. astaci* (Figure 5.9 B), *A. frigidophilus* isolates (Figure 5.9 C) and *S. parasitica* (Figure 5.9 E). Figure 3.23 shows the results of the restriction digestion assays on the PCR products, in particular Figure 3.23 A shows that enzymatic digestion only worked on the PCR product in well number 11, which is *A. astaci* KV (genotype C), while it did not occurr on isolates belonging to other genotypes. Figure 3.23 B and C show that the digestion patterns of *A. frigidophilus* AP5 and *S. parasitica* differs from the digestion pattern of *A. astaci* KV (genotype C).

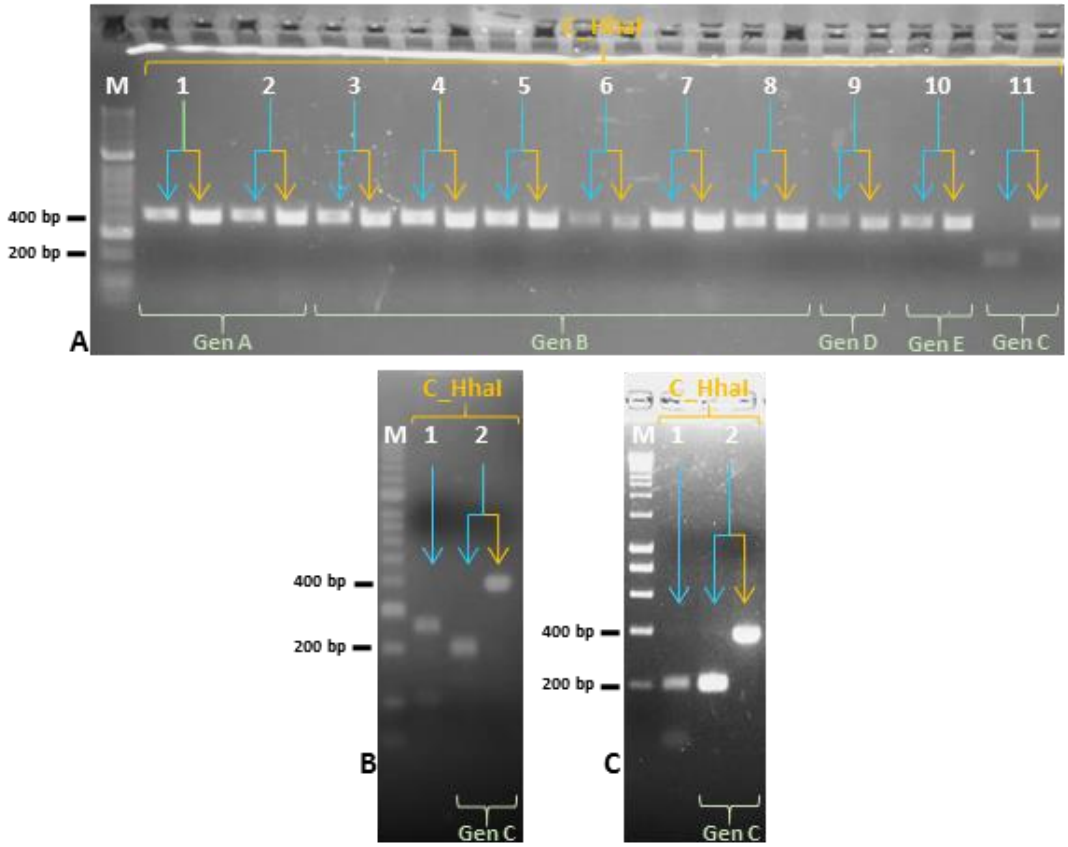


Figure 3.23 Restriction digestions of PCR products from *Aphanomyces* and oomycetes isolates amplified with primer pair C_HhaI. **A**: M, Bioline HyperLadder II; 1, *A. astaci* SV (genotype A); 2, *A. astaci* Da (genotype A); 3, *A. astaci* SA (genotype B); 4, *A. astaci* 197901 (genotype B); 5, *A. astaci* 457 (genotype B); 6, *A. astaci* Si (genotype B); 7, *A. astaci* D2 (genotype B); 8, *A. astaci* YX (genotype B); 9, *A. astaci* Pc (genotype D); 10, *A. astaci* KB13 (genotype E); 11, *A. astaci* KV (genotype C). **B**: M, Bioline HyperLadder II; 1, *A. frigidophilus* AP5; 2, *A. astaci* KV (genotype C). **C**: M, Bioline HyperLadder I; 1, *S. parasitica*; 2, *A. astaci* KV (genotype C). Blue arrows: PCR product after enzymatic digestion. Yellow arrows: original PCR product.

The primer pair D_HhaI amplified DNA regions of all *A. astaci* and *A. frigidophilus* isolates (Figure 5.10 A and C). No amplification was obtained from other *Aphanomyces* and oomycetes tested (Figure 5.10 C and E). Figure 3.24 shows the results of the restriction digestion assays on the PCR products. In particular Figure 3.24 A shows that enzymatic digestion only had an effect on the PCR product in well number 11, which is *A. astaci* Pc (genotype D), while it did not occurr on isolates belonging to other genotypes. Figure 3.24 B shows that the digestion patterns of *A. frigidophilus* isolates are different from the from the digestion pattern of *A. astaci* Pc (genotype D).

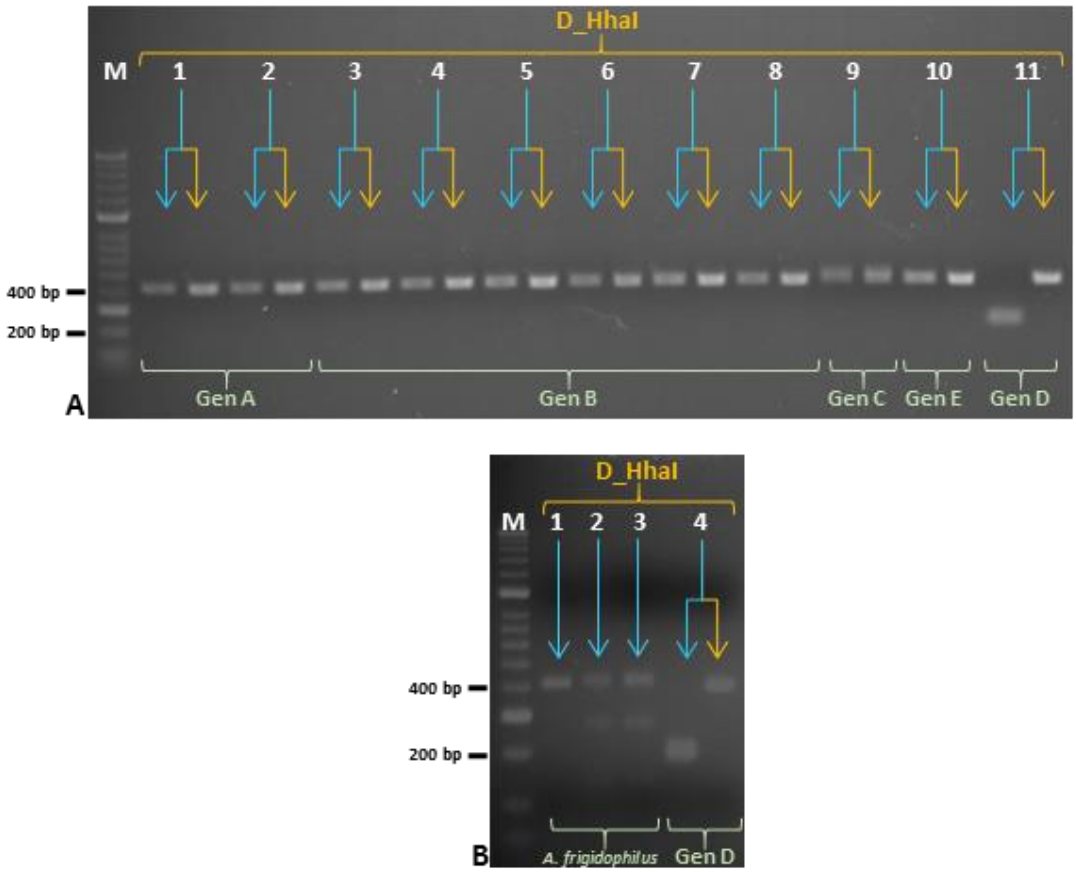


Figure 3.24 Restriction digestions of PCR products from *Aphanomyces* isolates amplified with primer pair D_HhaI. **A**: 1, *A. astaci* SV (genotype A); 2, *A. astaci* Da (genotype A); 3, *A. astaci* SA (genotype B); 4, *A. astaci* 197901 (genotype B); 5, *A. astaci* 457 (genotype B); 6, *A. astaci* Si (genotype B); 7, *A. astaci* D2 (genotype B); 8, *A. astaci* YX (genotype B); 9, *A. astaci* KV (genotype C); 10, *A. astaci* KB13 (genotype E); 11, *A. astaci* Pc (genotype D). **B**: 1, *A. frigidophilus* AP5; 2, *A. frigidophilus* RP1; 3, *A. frigidophilus* RP2; 4, *A. astaci* Pc (genotype D). M, Bioline HyperLadder I. Blue arrows: PCR product after enzymatic digestion. Yellow arrows: original PCR product.

The primer pair E_HhaI amplified DNA regions of all *A. astaci*, *A. invadans* and *A. frigidophilus* isolates (Figure 5.10 B and D) and the oomycetes *S. parasitica, S. furcata, L. caudata* and *A. racemosa* (Figure 5.10 F). As the PCR products of *S. parasitica* and *S. furcata* were of different size in comparison to *A. astaci* KB13 (genotype E), the restriction digestion was not applied on their PCR products. Figure 3.25 shows the results of the restriction digestion assays on the PCR products from the other isolates. Figure 3.25 A shows that enzymatic digestion only affected the PCR product of well number 11, which is *A. astaci* KB13 (genotype E), while it did not affect isolates belonging to other genotypes. Figure 3.25 B and C show that the digestion patterns of *A. invadans, A. frigidophilus, L. caudata* and *A. racemosa* PCR products are different from the from the digestion pattern of *A. astaci* KB13 (genotype E). The digestion patterns of *A. frigidophilus* isolates, *L. caudata* and *A. racemosa* present extra bands in comparison to *A. Astaci* KB13 (genotype E). This might be caused by the presence of other restricion sites in the region amplified by the primers and cut by the enzyme HhaI.

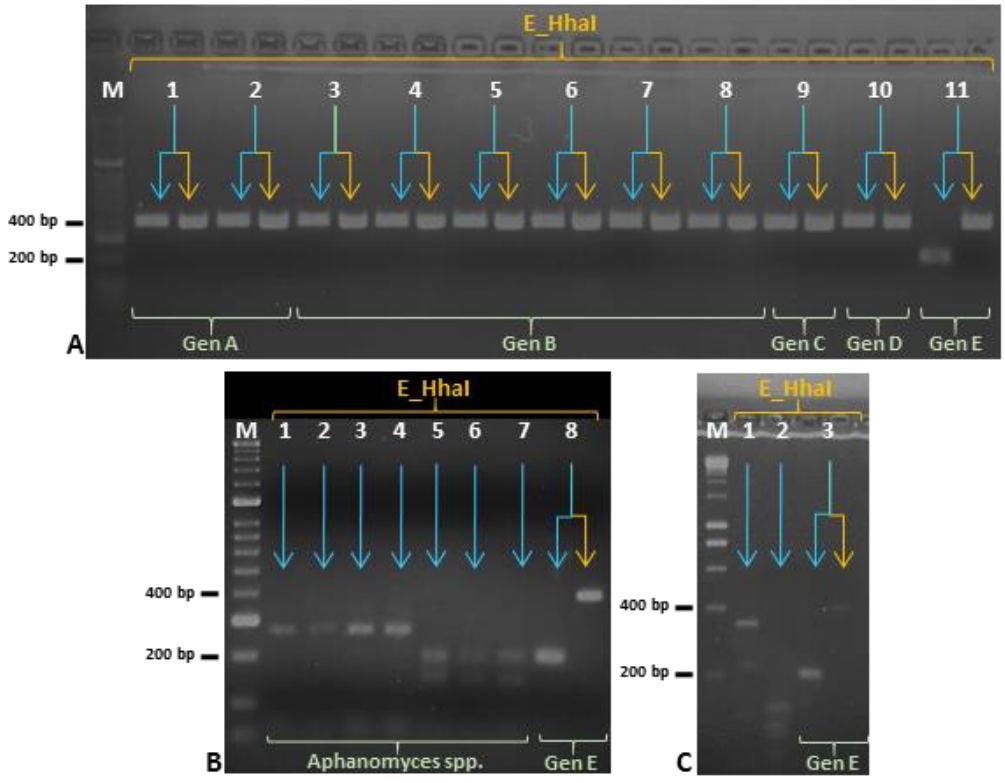


Figure 3.25 Restriction digestions of PCR products from *Aphanomyces* and oomycetes isolates amplified with primer pair E_HhaI. **A**: M, Bioline HyperLadder II; 1, *A. astaci* SV (genotype A); 2, *A. astaci* Da (genotype A); 3, *A. astaci* SA (genotype B); 4, *A. astaci* 197901 (genotype B); 5, *A. astaci* 457 (genotype B); 6, *A. astaci* Si (genotype B); 7, *A. astaci* D2 (genotype B); 8, *A. astaci* YX (genotype B); 9, *A. astaci* KV (genotype C); 10, *A. astaci* Pc (genotype D); 11, *A. astaci* KB13 (genotype E). **B**: M, Bioline HyperLadder II; 1, *A. invadans* NJM9030; 2, *A. invadans* NJM8997; 3, *A. invadans* NJM9701; 4, *A. invadans* NJM0002; 5, *A. frigidophilus* AP5; 6, *A. frigidophilus* RP1; 7, *A. frigidophilus* RP2; 8, *A. astaci* KB13 (genotype E). **C**: M, HyperLadder I; 1, *L. caudata*; 2, *A. racemosa*; 3, *A. astaci* KB13 (genotype E). Blue arrows: PCR product after enzymatic digestion. Yellow arrows: original PCR product.

To qualitatively confirm the presence/absence of the restriction site, PCR products of at least one *A. astaci* isolate for each genotype obtained by amplification with all five sets of primers were sequenced and analysed. As expected, the sequences reflected the *in silico* assays (Figure 5.12).

The restriction digestion assays were also applied to samples from outbreak 28325 (section 2.1.8). Sample 1.2 was tested with primer pairs B_HhaI, C_HhaI, D_HhaI and E_HhaI, while primer pair A_HhaI was used on samples 1.2 and 1.4 from the same outbreak. As the technique was still in development, all primer pairs were not tested on all available samples to preserve the samples to future testing. In the PCRs involving primer pairs A_HhaI, B_HhaI, C_HhaI and D_HhaI, the primers successfully amplified the regions selected and the PCR products were visible on agarose gel (Figure 5.11 A and B). Instead, no amplification was achieved with E_HhaI primer pair (Figure 5.11 B). PCR products obtained from A_HhaI, B_HhaI, C_HhaI and D_HhaI PCRs were then digested and results are shown in Figure 3.26. From these restriction digestions, the *A. astaci* present in outbreak 28325 sample 1.2 can be ascribed to genotype B as the digestion pattern of the sample (Figure 3.26 B, well number 1) matches the one of *A. astaci* D2 (genotype B) (Figure 3.26 B, well number 2), while it has not been digested in the other restriction assays.

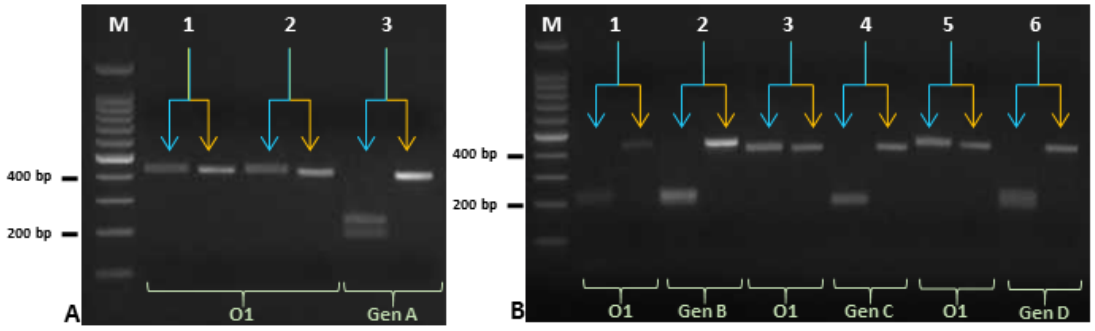


Figure 3.26 Restriction digestion of PCR products from outbreak PM28325 (O1). Primer-pairs used for PCR were B_HhaI, C_HhaII and D_HhaI. **A**: 1, O1-1.2; 2, O1-1.4; 3, *A. astaci* Da (genotype A). **B**: 1, O1-1.2; 2, *A. astaci* D2 (genotype B); 3, O1-1.2; 4, *A. astaci* KV (genotype C); 5, O1-1.2; 6, *A. astaci* Pc (genotype D). M, Promega 100 bp DNA Ladder. Blue arrows: PCR product after enzymatic digestion. Yellow arrows: original PCR product.

To improve the amplification of *A. astaci* present in outbreak 28325 sample 1.2 with E_HhaI primer pair, a touchdown PCR (section 2.3.4.7) using increased cycle numbers was applied without success (results not shown).

The primer pairs were also tested on Cefas carriers' samples (section 2.1.8), (*i.e.* gDNA extractions from North American crayfish tissues known to be infected with *A. astaci* but asymptomatic) Due to likely low concentration of *A. astaci* gDNA, amplification of the targeted regions was not possible and PCRs were negative (results not shown).

3.4.4 Development of restriction digestion assays to detect and discriminate *Aphanomyces astaci* genotypes A, C, D, and E targeting mitochondrial DNA

The restriction digestion assays developed in this study to detect and distinguish *A. astaci* genotypes was successfully applied to genomic DNA from pure cultures and an outbreak samples. However, an increase in sensitivity is required to be able to detect *A. astaci* gDNA from carrier samples. As eukaryotic cells contain a high number of mitochondria (and therefore copies of mitochondrial genomic DNA) per nucleus (and therefore nuclear genomic DNA) (Luo *et al.*, 2011; Waugh, 2007), by targeting mitochondrial DNA sensitivity may be improved. Therefore, a similar bioinformatics approach was used to detect genotype-specific SNVs and develop restriction digestion assays targeting the mitochondrial DNA. In order to identify high-confidence inter-genotype polymorphisms, the genomic positions of an *A. astaci* newly assembled mitochondrial genome were categorised as ambiguous or unambiguous and for each *A. astaci* genotype, genotype-specific SNVs in restriction sites were identified, visualised in IGV (genotypes A, C, D and E: Figure 3.27, yellow boxes) and primers to amplify the fragments were designed. Table 3.5 lists the numbers of genotype-specific SNVs found in the mitochondrial genome of each genotype. The number of SNVs in the mtDNA is considerably lower in comparison to the nuclear DNA (Table 3.1), results expected as the mtDNA is smaller than the nuclear DNA.

Table 3.5 Numbers of mitochondrial genotype specific SNVs. Each isolate belonging to a genotype shares these SNVs with the isolate belonging to the same genotype.

Genotype	SNVs
A	6
B	4
C	2
D	92
E	4

For genotype B, the SNVs detected were not in restriction sites (Figure 3.32, yellow box), therefore one of the variants was selected and a semi-nested PCR developed (sections 2.3.4.9 and 3.4.5). Figure 3.27 displays the alignments of the 12 *A. astaci* paired-end reads to *A. astaci* mitochondrion reference assembly. As before, each individual vertical bar represents the coverage track, which displays the depth of read coverage for each genomic position for each

isolate against a reference genome. If the bar is grey, most nucleobases aligned in that position to the reference genome are matching the reference genome while if the bar is of a different colour, the nucleobases aligned in that position do not match the reference genome. Green refers to the nucleobase variant adenosine (A), red to tyrosine (T), yellow to guanosine (G) and blue to cytosine (C).

For genotype A (Figure 3.27, top left), the mitochondrial genotype-specific SNV chosen is situated in the restriction site “CCGG”, cut by the restriction enzyme MspI. For genotype C (Figure 3.27, top right), the mitochondrial genotype-specific SNV chosen is situated in the restriction site “TTAA”, cut by the restriction enzyme MseI. For genotype D (Figure 3.27, bottom left), the mitochondrial genotype-specific SNV chosen is situated in the restriction site “CCGC”, cut by the restriction enzyme AciI. For genotype E (Figure 3.27, bottom right), the mitochondrial genotype-specific SNV chosen is situated in the site “TAGA”, a site that is not cut by any known restriction enzymes. For this last SNV, the isolates not belonging to genotype E present the variant “TCGA” that is cut by the enzyme Taqᵅ I.

The primers to amplify the regions containing the SNVs were designed at ~250 bp before and after the SNVs (final fragment of ~500 bp) to facilitate visualisation of PCR products on agarose gels after amplification and after digestion.

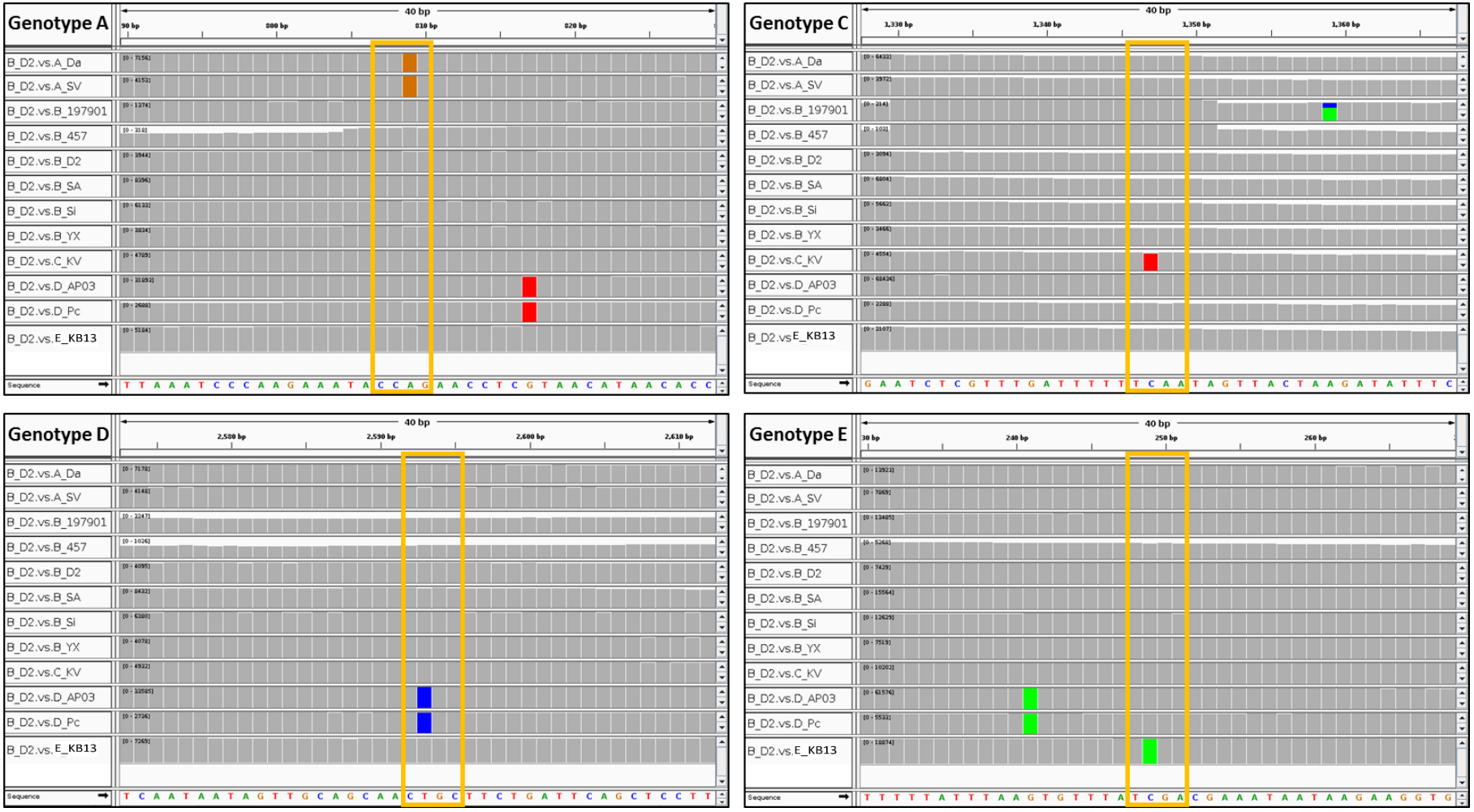


Figure 3.27 Mitochondrial SNVs unique to *A. astaci* genotypes A, C, D, and E that fall in restriction sites. Alignments of genomic sequence reads against a reference genome visualised in IGV. Yellow boxes: SNVs in restriction sites. Alignments of *A. astaci* isolates paired-end reads from top to bottom: Da (genotype A); SV (genotype A); 197901 (genotype B); 457 (genotype B); D2 (genotype B); SA (genotype B); Si (genotype B); YX (genotype B); KV (genotype C); APO3 (genotype D); Pc (genotype D); KB13 (genotype E). Reference genome "Sequence": *A. astaci* D2 (genotype B) mitochondrion. Individual vertical bars: depth of read coverage of each genomic position against the reference genome. Grey bars: majority (> 80 %) of nucleobases aligned match the reference genome. Green, red, yellow, and blue bars: at least 20 % of nucleobases aligned not matching the reference genome and substituted respectively by adenosine (A), tyrosine (T), guanosine (G) and cytosine (C).

The primer pairs were first tested on diluted *A. astaci* gDNA (Table 5.11) to test whether the sensitivity of the test was increased and a lower concentration of DNA could indeed be detected. Direct comparisons between the PCR products after a 35 cycles PCR amplification with primers targeting the mtDNA (Figure 5.13) and the nuclear DNA (Figure 5.14) shows that sensitivity of the former is higher. The PCR followed by restriction digestion assays were then tested on the other *Aphanomyces* species and oomycetes available in Cefas OCC (Table 2.1 and Table 2.3).

The primer pair A_MITO amplified DNA regions in all *Aphanomyces* isolates (results not shown) and the oomycetes *S. parasitica, S. furcata, L. caudata, A. racemosa*, *A. laevis* and *P. monospermum* (Figure 5.15 A). Figure 3.28 shows the results of the restriction digestion assay on the PCR products obtained from the PCR reactions. The enzymatic digestion occurred only for the PCR product in wells number 1 and 20 (Figure 3.28 A) and 7 (Figure 3.28 B), which are respectively *A. astaci* SV and Da (genotype A), while it did not occur for the other isolates.

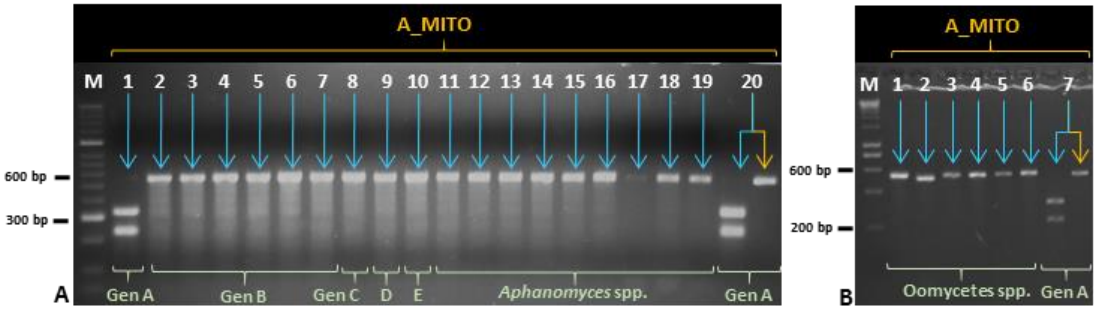


Figure 3.28 Enzymatic digestions of PCR products from *Aphanomyces* and oomycetes isolates amplified with primer pair A_MITO. **A**: M, Bioline HyperLadder II; 1, *A. astaci* SV (genotype A); 2, *A. astaci* D2 (genotype B); 3, *A. astaci* YX (genotype B); 4, *A. astaci* 197901 (genotype B); 5, *A. astaci* 457 (genotype B); 6, *A. astaci* Si (genotype B); 7, *A. astaci* SA (genotype B); 8, *A. astaci* KV (genotype C); 9, *A. astaci* Pc (genotype D); 10, *A. astaci* KB13 (genotype E); 11, *A. invadans* NJM8997; 12, *A. invadans* NJM9030; 13, *A. invadans* NJM9701; 14, *A. invadans* NJM0002; 15, *A. invadans* GWR; 16, *A. invadans*-like NJM9510; 17, *A. frigidophilus* AP5; 18, *A. frigidophilus* RP1; 19, *A. frigidophilus* RP2; 20, *A. astaci* Da (genotype A). **B**: Bioline HyperLadder I; 1, *S. parasitica*; 2, *S. furcata*; 3, *L. caudata*; 4, *A. racemosa*; 5, *A. laevis*; 6, *P. monospermum*; 7, *A. astaci* Da (genotype A). Blue arrows: PCR product after enzymatic digestion. Yellow arrows: original PCR product.

The primer pair C_MITO amplified DNA regions in all *Aphanomyces* and oomycetes isolates tested (results not shown). Figure 3.29 shows the results of restriction digestion assay on the PCR products. The enzymatic digestion occurred for all PCR fragments but the pattern of the bands on the gels are different from the ones produced by *A. astaci* KV (genotype C) (Figure 3.29 A, well 20 and B, well 9).

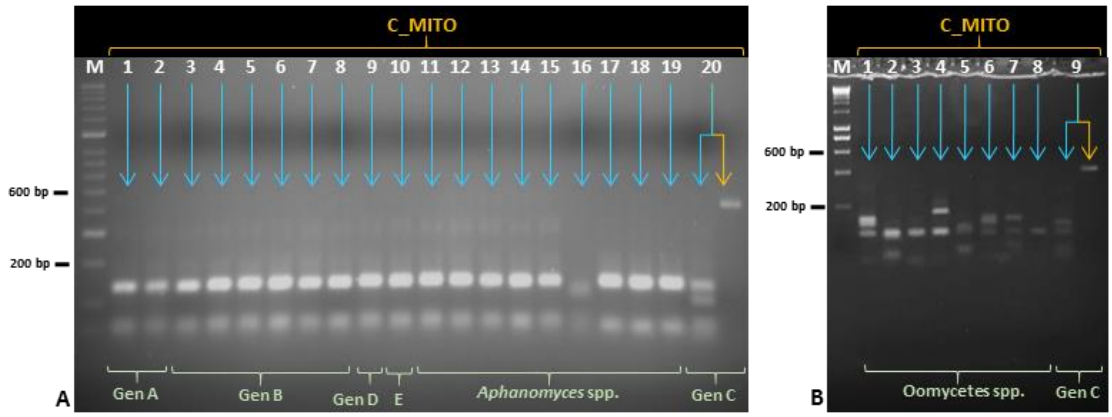


Figure 3.29 Enzymatic digestions of PCR products from *Aphanomyces* and oomycetes isolates amplified with primer pair C_MITO. **A**: M, Bioline HyperLadder II; 1, *A. astaci* SV (genotype A); 2, *A. astaci* Da (genotype A); 3, *A. astaci* D2 (genotype B); 4, *A. astaci* YX (genotype B); 5, *A. astaci* 197901 (genotype B); 6, *A. astaci* 457 (genotype B); 7, *A. astaci* Si (genotype B); 8, *A. astaci* SA (genotype B); 9, *A. astaci* Pc (genotype D); 10, *A. astaci* KB13 (genotype E); 11, *A. invadans* NJM8997; 12, *A. invadans* NJM9030; 13, *A. invadans* NJM9701; 14, *A. invadans* NJM0002; 15, *A. invadans* GWR; 16, *A. invadans*-like NJM9510; 17, *A. frigidophilus* AP5; 18, *A. frigidophilus* RP1; 19, *A. frigidophilus* RP2; 20, *A. astaci* KV (genotype C). **B**: Bioline HyperLadder I; 1, *S. parasitica*; 2, *S. furcata*; 3, *L. caudata*; 4, *A. racemosa*; 5, *A. laevis*; 6, *P. monospermum*; 7, *P. flevoense*; 8, *Phoma*-like; 9, *A. astaci* KV (genotype C). Blue arrows: PCR product after enzymatic digestion. Yellow arrows: original PCR product.

The primer pair D_MITO amplified DNA regions in all *Aphanomyces* and oomycetes isolates tested (results not shown). Figure 3.30 shows the results of the restriction digestion assay on the PCR products. While the enzymatic digestion of *A. invadans*-like NJM9510 (Figure 3.30 A, well 16) and *S. furcata* (Figure 3.30 B, well 2) PCR products produced different pattern of bands from the one of *A. astaci* Pc (genotype D) (Figure 3.30 A, well 20 and B, well 9), *L. caudata* and *Phoma*-like (Figure 3.30 B, wells 3 and 8) show similar digestion patterns. The PCR products of *L. caudata* and *Phoma*-like amplified with D_MITO primers were purified and sequenced in order to investigate the similarities/differences with *A. astaci* Pc sequences. The alignment of these fragments revealed the presence of the SNV in the restriction site for *A. astaci* Pc (genotype D), *L. caudata* and *Phoma*-like isolates but overall sequence difference between the isolates (Figure 5.16).

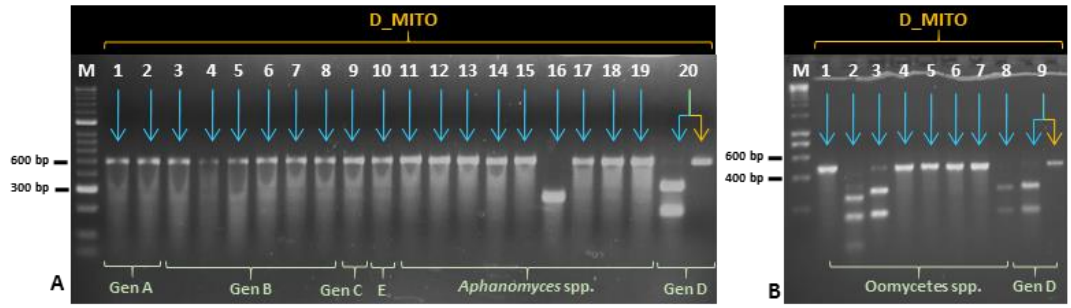


Figure 3.30 Enzymatic digestions of PCR products from *Aphanomyces* and oomycetes isolates amplified with primer pair D_MITO. **A**: M, Bioline HyperLadder II; 1, *A. astaci* SV (genotype A); 2, *A. astaci* Da (genotype A); 3, *A. astaci* D2 (genotype B); 4, *A. astaci* YX (genotype B); 5, *A. astaci* 197901 (genotype B); 6, *A. astaci* 457 (genotype B); 7, *A. astaci* Si (genotype B); 8, *A. astaci* SA (genotype B); 9, *A. astaci* KV (genotype C); 10, *A. astaci* KB13 (genotype E); 11, *A. invadans* NJM8997; 12, *A. invadans* NJM9030; 13, *A. invadans* NJM9701; 14, *A. invadans* NJM0002; 15, *A. invadans* GWR; 16, *A. invadans*-like NJM9510; 17, *A. frigidophilus* AP5; 18, *A. frigidophilus* RP1; 19, *A. frigidophilus* RP2; 20, *A. astaci* Pc (genotype D). **B**: Bioline HyperLadder I; 1, *S. parasitica*; 2, *S. furcata*; 3, *L. caudata*; 4, *A. racemosa*; 5, *A. laevis*; 6, *P. monospermum*; 7, *P. flevoense*; 8, *Phoma*-like; 9, *A. astaci* Pc (genotype D). Blue arrows: PCR product after enzymatic digestion. Yellow arrows: original PCR product.

The primer pair E_MITO amplified DNA regions in all *A. astaci, A. frigidophilus* and *A. invadans* NJM9701 and GWR (Figure 5.15 C) and the oomycetes *S. furcata* and *A. laevis* (Figure 5.15 B). Figure 3.31 shows the results of the restriction digestion assay on the PCR products. The enzymatic digestion occurred for all PCR fragments but the pattern of the bands on the gels are different from the ones produced by *A. astaci* KB13 (genotype E) (Figure 3.31 A, well 16 and B, well 3).

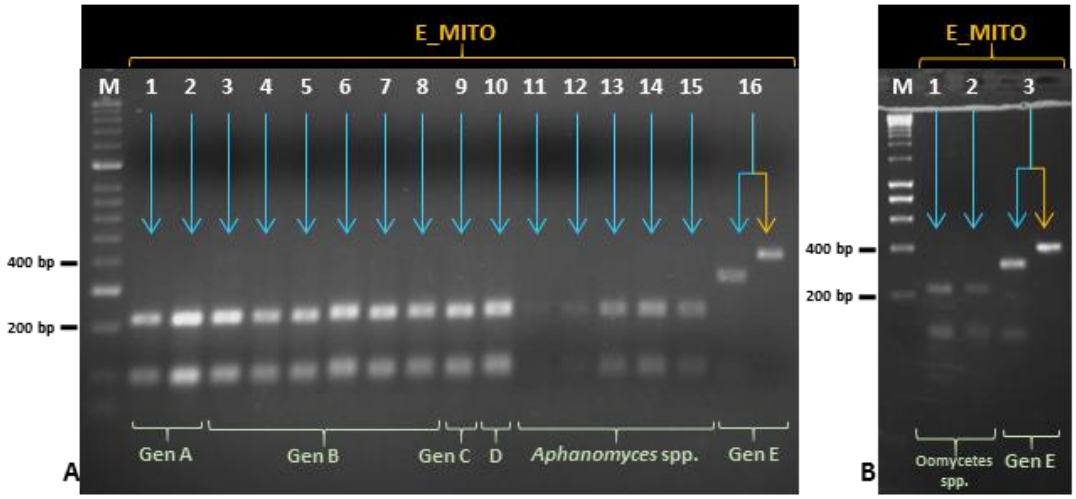


Figure 3.31 Enzymatic digestions of PCR products from *Aphanomyces* and oomycetes isolates amplified with primer pair E_MITO. **A**: M, Bioline HyperLadder II; 1, *A. astaci* SV (genotype A); 2, *A. astaci* Da (genotype A); 3, *A. astaci* D2 (genotype B); 4, *A. astaci* YX (genotype B); 5, *A. astaci* 197901 (genotype B); 6, *A. astaci* 457 (genotype B); 7, *A. astaci* Si (genotype B); 8, *A. astaci* SA (genotype B); 9, *A. astaci* KV (genotype C); 10, *A. astaci* Pc (genotype D); 11, *A. invadans* NJM9701; 12, *A. invadans* GWR; 13, *A. frigidophilus* AP5; 14, *A. frigidophilus* RP1, 15, *A. frigidophilus* RP2; 16, *A. astaci* KB13 (genotype E). **B**: Bioline HyperLadder I; 1, *S. furcata*; 2, *A. laevis*; 3, *A. astaci* KB13 (genotype E). Blue arrows: PCR product after enzymatic digestion. Yellow arrows: original PCR product.

To qualitatively confirm the presence/absence of the SNVs in the restriction site, PCR products of at least one *A. astaci* isolate for each genotype obtained by amplification with A_MITO, C_MITO, D_MITO and E_MITO primer sets were sequenced and analysed. As expected, the sequences reflected the *in silico* assays (Figure 5.17).

As the restriction digestion assays here developed were able to distinguish genotypes A, C, D, and E, the primers and assays were tested on Cefas historical samples (sections 3.4.6 and 3.4.7).

3.4.5 Development of a semi-nested assay to detect and discriminate *Aphanomyces astaci* genotype B targeting mitochondrial DNA

The restriction digestion assays developed in this study to detect and distinguish *A. astaci* genotypes A, C, D, and E were successful and the sensitivity of the assays was tested on pure cultures in section 3.4.4. As it was not possible to identify SNVs unique to genotype B within restriction sites, one of the variants was selected (Figure 3.32, yellow box) and a semi-nested PCR developed (2.3.4.8) for a digest-free assay.

Figure 3.32 display the alignments of the 12 *A. astaci* paired-end reads to *A. astaci* mitochondrion reference assembly and for genotype B isolates (197901, 457, D2, SA, Si, and YX). The variant selected presents the nucleotide “G” while the other isolates present the variant “A”.

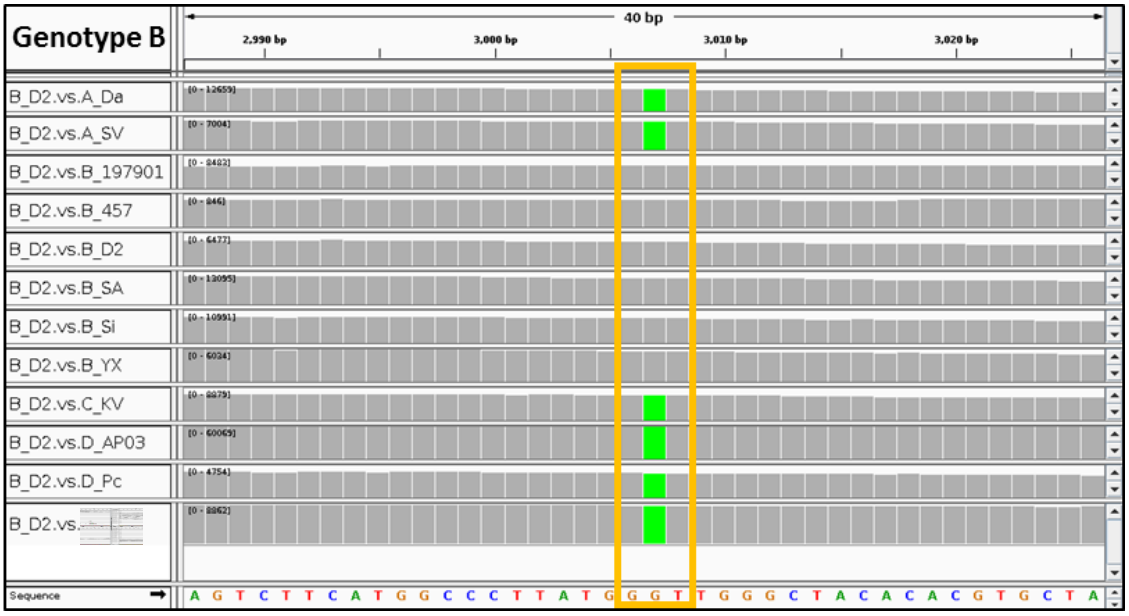


Figure 3.32 Mitochondrial SNV unique to *A. astaci* genotype B. Alignments of sequence reads against a reference genome visualised in IGV. Yellow box: SNV. Alignments of *A. astaci* isolates paired-end reads from top to bottom: Da (genotype A); SV (genotype A); 197901 (genotype B); 457 (genotype B); D2 (genotype B); SA (genotype B); Si (genotype B); YX (genotype B); KV (genotype C); APO3 (genotype D); Pc (genotype D); KB13 (genotype E). Reference genome “Sequence”: *A. astaci* D2 (genotype B) mitochondrion. Individual vertical bars: depth of read coverage of each genomic position against the reference genome. Grey bars: majority (> 80 %) of nucleobases aligned match the reference genome. Green bars: at least 20 % of nucleobases aligned not matching the reference genome and substituted by adenosine (A).

As for genotype-specific SNVs for genotype A, C, D, and E, the primers to amplify the region containing genotype B SNV were designed at ~250 bp before and after the SNV (final fragment of ~500 bp) to facilitate visualisation of PCR products on agarose gels after the first round PCR.

As described in section 2.3.4.9, a semi-nested PCR was developed to exploit the SNV present in the region amplified by B_MITO primers. Two primers of equal size and nucleotide sequence, except for the last one, were designed:

- B_MITO_N1S: amplification of mitochondrial region with last base (*i.e.* 3') matching the genotype B SNV;
- B_MITO_N2S: amplification of mitochondrial region with last base (*i.e.* 3') not matching the genotype B SNV, used as PCR control primer.

Amplifying the PCR product obtained from the first round PCR (following protocol section 2.4.3.7, primer pair B_MITO) with the semi-nested primers, the presence or absence of a second band of smaller size in comparison to the original PCR product on the agarose gel indicates that the primers matched the variant in the template (Chen and Sullivan, 2003). By using one nested primer at a time in the PCR reaction, the PCR product can be ascribed to genotype B or not. The sensitivity of primer pair B_MITO was tested on an *A. astaci* gDNA dilution (Table 5.11) and comparison of PCR products after a 35 cycles amplification with primers targeting mtDNA (Figure 5.13) and nuclear DNA (Figure 5.14) shown that sensitivity of the former is higher.

To understand if the semi-nested PCR developed here could distinguish genotype B from other genotypes, a first round PCR with B_MITO primer pair was applied to one *A. astaci* isolate for each genotype, followed by semi-nested PCR. As shown in Figure 3.33, the presence of a band around the size of 200 bp indicates the amplification of half of the first-round PCR product. For *A. astaci* D2 (genotype B), amplification was successful only in association with the primer B_MITO_N1S (well 2), while no amplification occurred with the primer B_MITO_N2S (well 8). Conversely, PCR products from *A. astaci* belonging to the other genotypes were not amplified by the primer B_MITO_N1S (wells: 1 and 3-5) while successfully amplified by the primer B_MITO_N2S (wells: 7 and 9-11).

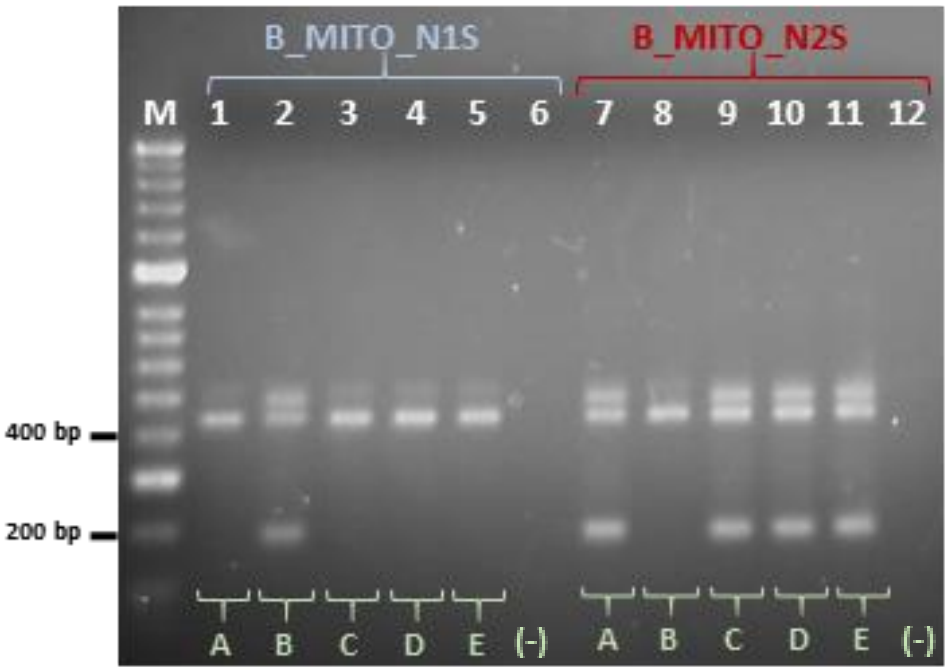


Figure 3.33 Semi-nested PCR on *A. astaci* isolates. M, Bioline HyperLadder II; 1-6: B_MITO_N1S primer: 1, *A. astaci* Da (genotype A); 2, *A. astaci* D2 (genotype B); 3, *A. astaci* KV (genotype C); 4, *A. astaci* Pc (genotype D); 5, *A. astaci* KB13 (genotype E); 6, negative control. 7-12: B_MITO_N2S primer: 7, *A. astaci* Da (genotype A); 8, *A. astaci* D2 (genotype B); 9, *A. astaci* KV (genotype C); 10, *A. astaci* Pc (genotype D); 11 *A. astaci* KB13 (genotype E); 12, negative control. (-), negative control.

As the semi-nested PCR proved to be a useful tool to distinguish genotype B from the other genotypes, the same approach was applied to the other *A. astaci* and *Aphanomyces* and oomycetes isolates (Table 2.1 and Table 2.3). All *Aphanomyces* and oomycetes isolated tested positive at the first round PCR (results not shown) and PCR products were submitted to semi-nested PCR.

A. astaci D2 (genotype B) was used as positive control and *A. astaci* Da (genotype A) as negative control for the semi-nested PCR with primer B_MITO_N1S, while *A. astaci* D2 (genotype B) was used as negative control and *A. astaci* Da (genotype A) as positive control for the semi-nested PCR with primer B_MITO_N2S. A negative control with no DNA template was used in each reaction to detect contaminations. Figure 3.34 A shows the results of the semi-nested PCRs with primer B_MITO_N1S, which demonstrates that the amplification occurred only in *A. astaci* isolates belonging to genotype B (wells: 2-6 and 17) and no interference was displayed in any other *Aphanomyces* species. From Figure 3.34 B, instead, the primer B_MITO_N2S amplified all *Aphanomyces* isolates except for *A. astaci* belonging to genotype B (wells: 2-6 and 17).

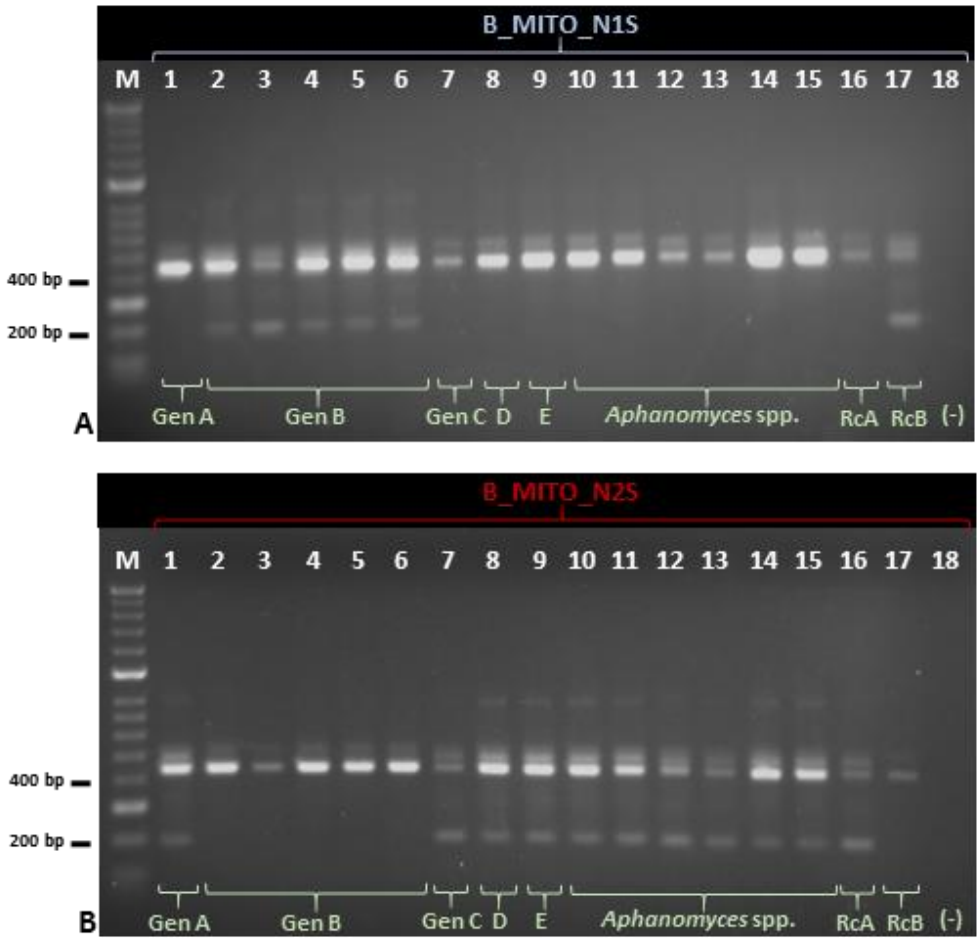


Figure 3.34 Semi-nested PCR on *Aphanomyces* isolates (Table 2.1). **A**: B_MITO_N1S primer. **B**: B_MITO_N2S primer. In both: M, Bioline HyperLadder II; 1, *A. astaci* SV (genotype A); 2, *A. astaci* 457 (genotype B); 3, *A. astaci* 197901 (genotype B); 4, *A. astaci* Si (genotype B); 5, *A. astaci* SA (genotype B); 6, *A. astaci* YX (genotype B); 7, *A. invandas* NJM9030; 8, *A. invandas* NJM9701; 9, *A. invandas* NJM0002; 10, *A. invandas*, GWR; 11, *A. invandas* NJM8997; 12, *A. invadans*-like NJM9510; 13, *A. frigidophilus* AP5; 14, *A. frigidophilus* RP1; 15, *A. frigidophilus* RP2; 16, *A. astaci* Da (genotype A); 17, *A. astaci* D2 (genotype B); 18, negative control. RcA, PCR reaction control (genotype A); RcB, PCR reaction control (genotype B); (-), negative control.

Similar results were obtained applying semi-nested PCR on oomycetes isolates with B_MITO_N1S primer: amplification occurred only in *A. astaci* D2 (genotype B) used as positive control (Figure 3.35, well 9). For B_MITO_N2S primer, amplification was obtained only in *A. laevis* and *A. astaci* Da (genotype A) used as positive control (Figure 3.35, wells 5 and 10). As the semi-nested PCR reactions worked for both primers (testified by the positive controls in wells 9A and 10B), the lack of PCR bands in wells 1-8A, 1-4B and 6-8B means that these oomycetes isolates don't share the same sequence at either the B_MITO_N1S and B_MITO_N2S primers sites.

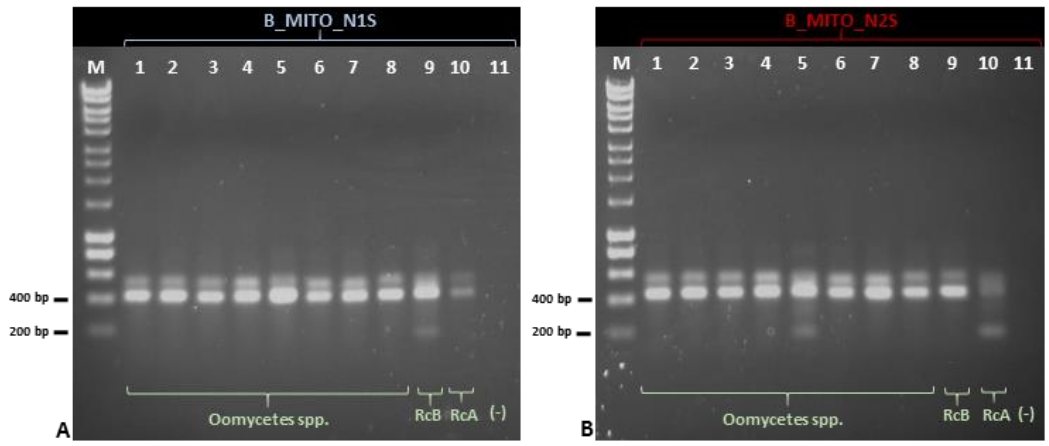


Figure 3.35 Semi-nested PCR on oomycete isolates. **A**: B_MITO_N1S primer. **B**: B_MITO_N2S primer. In both: M, Bioline HyperLadder I; 1, *S. parasitica*; 2, *S. furcata*; 3, *L. caudata*; 4, *A. racemosa*; 4, *A. laevis*; 5, *P. monospermum*; 6, *P. flevoense*; 7, *Phoma*-like; 9, *A. astaci* D2 (genotype B); 10, *A. astaci* Da (genotype A); 11, negative control. RcA, PCR reaction control (genotype A); RcB, PCR reaction control (genotype B); (-), negative control.

To qualitatively confirm the presence/absence of the selected SNV, PCR products of at least one *A. astaci* isolate for each genotype obtained by amplification with B_MITO primer pair were sequenced and analysed. As expected, the sequences reflected the *in silico* assays (Figure 5.17).

As the semi-nested PCR developed here could distinguish genotype B isolates from other *Aphanomyces* and oomycetes species, the assay was subsequently tested on Cefas historical samples (2.1.8).

3.4.6 Restriction digestion and semi-nested assays on historic crayfish plague outbreaks samples

The restriction digestion assays and the semi-nested PCR were tested on Cefas crayfish outbreaks historical samples (section 2.1.8). As previously mentioned, these samples suffer from low amounts of *A. astaci* gDNA in comparison to pure cultures where ample amounts of gDNA can be extracted. Primer pairs A_MITO, B_ MITO, C_ MITO, D_ MITO, and E_ MITO were used to amplify the targeted regions from the outbreak samples (Table 5.2) and amplification was achieved in the majority of samples tested (Figure 5.18 A and B; Figure 5.19 A, B, C, D and E). Table 5.12 summarises the results of the PCR reactions on the outbreaks samples.

PCR products obtained with primer pairs A_MITO, C_ MITO, D_ MITO and E_ MITO were subjected to restriction digestion assays (Figure 5.18 C and D; Figure 5.20 A, B and C), while PCR products obtained with primers pairs B_MITO were subjected to semi-nested PCRs (Figure 5.21 A and B).

Table 3.6 summarises the results of the restriction digestions and the semi-nested PCRs.

All samples, from which a PCR product was obtained, were negative for the presence of genotypes C, D, and E, but outbreak PM21018 sample number 2 was positive for *A. astaci* genotype A. All outbreaks tested positive for the presence of *A. astaci* genotype B, including the sample number 1 and 3 from outbreak PM21018, indicating the simultaneously presence of two different genotypes in the outbreak. Aligning the sequenced PCR product of PM21018 sample number 2 (amplified with A_MITO primer pairs) to *A. astaci* Da (genotype A) clearly demonstrate the identical nature of the sequences and the presence of the SNV in the restriction site, therefore confirming the double nature of the outbreak (Figure 5.22).

Table 3.6 Results of restriction digestion assays and semi-nested PCR for products amplified by A, B, C, D, and E mitochondrial primer pairs from outbreak samples. N/A: sample negative to previous PCR reaction; +: sample positive for restriction digestion or semi-nested PCR; -: sample negative at restriction digestion or semi-nested PCR assays; yellow: successful detection and distinction of genotype; light blue: successful detection and distinction of non-genotype.

9	Sample	A_MITO Restriction digestion		B_MITO_N1S Semi-nested		B_MITO_N2S Semi-nested		C_MITO Restriction digestion		D_MITO Restriction digestion		E_MITO Restriction digestion	
PM28325	**1.2**	-	Figure 5.18 C well 1	+	Figure 5.21 A well 1	-	Figure 5.21 A well 1	-	Figure 5.18 C well 4	-	Figure 5.18 C well 8	-	Figure 5.18 D well 1
	1.4	-	Figure 5.18 C well 2	+	Figure 5.21 A well 2	-	Figure 5.21 A well 2	-	Figure 5.18 C well 5	-	Figure 5.18 C well 9	-	Figure 5.18 D well 2
PM28465	**1.2**	N/A		+	Figure 5.21 A well 3	-	Figure 5.21 A well 3	-	Figure 5.18 C well 6	N/A		-	Figure 5.18 D well 3
PM21018	**1**	N/A		+	Figure 5.21 A well 4	-	Figure 5.21 A well 4	-	Figure 5.20 B well 1	-	Figure 5.20 A well 9	-	Figure 5.20 C well 1
	2	+	Figure 5.20 A well 1	N/A		N/A		-	Figure 5.20 B well 2	N/A		N/A	
	3	N/A		+	Figure 5.21 A well 5	-	Figure 5.21 A well 5	N/A		N/A		-	Figure 5.20 C well 2
PM19790	**1**	-	Figure 5.20 A well 2	+	Figure 5.21 A well 6	-	Figure 5.21 A well 6	-	Figure 5.20 B well 3	-	Figure 5.20 A well 10	-	Figure 5.20 C well 3
	2	-	Figure 5.20 A well 3	+	Figure 5.21 A well 7	-	Figure 5.21 A well 7	-	Figure 5.20 B well 4	-	Figure 5.20 A well 11	-	Figure 5.20 C well 4
	6	-	Figure 5.20 A well 4	+	Figure 5.21 A well 8	-	Figure 5.21 A well 8	-	Figure 5.20 B well 5	-	Figure 5.20 A well 12	-	Figure 5.20 C well 5
	7	-	Figure 5.20 A well 5	+	Figure 5.21 A well 9	-	Figure 5.21 A well 9	-	Figure 5.20 B well 6	-	Figure 5.20 A well 13	-	Figure 5.20 C well 6
PM19955	**1.1**	-	Figure 5.20 A well 6	+	Figure 5.21 A well 10	-	Figure 5.21 A well 10	-	Figure 5.20 B well 7	-	Figure 5.20 A well 14	-	Figure 5.20 C well 7
	2.1	N/A		N/A		N/A		-	Figure 5.20 B well 8			N/A	
	3.1	-	Figure 5.20 A well 7	+	Figure 5.21 A well 11	-	Figure 5.21 A well 11	-	Figure 5.20 B well 9	-	Figure 5.20 A well 15	-	Figure 5.20 C well 8

3.4.7 Restriction digestion and semi-nested assays on North American crayfish samples (carriers)

As for the outbreak samples, the restriction digestion assays and the semi-nested PCR were tested on Cefas crayfish carriers' historical samples (2.1.10) with primer pairs A_MITO, B_ MITO, C_ MITO, D_ MITO, and E_ MITO. Amplification of the targeted DNA region was achieved in the majority of samples tested (Figure 5.23 A; Figure 5.24 A; Figure 5.25 A; Figure 5.26 A; Figure 5.27 A) (Table 5.12) and PCR products obtained with primer pairs A_MITO, C_ MITO, D_ MITO, and E_ MITO were subjected to restriction digestion assays (Figure 5.23 B; Figure 5.25 B; Figure 5.26 B; Figure 5.27 B), while PCR products obtained with primer pairs B_MITO were subjected to semi-nested PCRs (Figure 5.24 B and C).

Table 3.7 summarises the results of the restriction digestions and the semi-nested PCRs. All carriers, from which a PCR product was obtained, were negative for the presence of genotypes A, C, D, and E and most of the samples resulted positive for the presence of *A. astaci* genotype B (Table 3.7). Three samples were both positive at the two B_MITO semi-nested primers. The respective first round PCR products were purified, sequenced and the sequences aligned to *A. astaci* genotype B and *A. astaci* mitochondrial reference sequence (genotype B). The SNV of sample III-28-C matched the ones from *A. astaci* D2 (genotype B) and *A. astaci* mitochondrial reference sequence (genotype B), the PCR product presents some differences in comparison to the reference sequence (Figure 5.28). The SNV of sample XVIII-16a1-C matched the ones from *A. astaci* D2 (genotype B) and *A. astaci* mitochondrial reference sequence (genotype B), the PCR product present a highly similar sequence to the reference sequence. The SNV of sample XVII-9a2-C does not match the one from *A. astaci* D2 (genotype B) and *A. astaci* mitochondrial reference sequence (genotype B), but the PCR product shared the same sequence (Figure 5.28). These alignments combined with the semi-nested PCR results indicate the presence of *A. astaci* genotype B and some other oomycete (or an unknown genotype) in the sample.

For some samples, it was not possible to detect PCR products with some of the primer pairs. However, failure to amplify something does not exclude its presence.

Table 3.7 Restriction digestion assays and semi-nested PCR results for PCR products amplified by A, B, C, D, and E mitochondrial primer pairs from carrier samples. N/A: sample negative to previous PCR reaction; +: sample positive for restriction digestion or semi-nested PCR; -: sample negative at restriction digestion or semi-nested PCR assays; yellow: successful detection and distinction of genotype; light blue: successful detection and distinction of non-genotype; orange: positive samples with ambiguous results.

Population	Sample N	A_MITO Restriction digestion		B_MITO_N1S Semi-nested PCR		B_MITO_N2S Semi-nested PCR		C_MITO Restriction digestion		D_MITO Restriction digestion		E_MITO Restriction digestion	
III	13-T	-	Figure 5.23 B well 1	+	Figure 5.24 B well 1	-	Figure 5.24 C well 1	-	Figure 5.25 B well 1	N/A		-	Figure 5.27 B well 1
III	28-C	N/A		+	Figure 5.24 B well 2	+	Figure 5.24 C well 2	-	Figure 5.25 B well 2	N/A		N/A	
IV	30-T	N/A		-	Figure 5.24 B well 3	-	Figure 5.24 C well 3	-	Figure 5.25 B well 3	-	Figure 5.26 B well 1	-	Figure 5.27 B well 2
VII	26-T	N/A		+	Figure 5.24 B well 4	-	Figure 5.24 C well 4	-	Figure 5.25 B well 4	N/A		N/A	
VIII	6-T	N/A		-	Figure 5.24 B well 5	-	Figure 5.24 C well 5	-	Figure 5.25 B well 5	N/A		N/A	
IX	6-C	-	Figure 5.23 B well 2	+	Figure 5.24 B well 6	-	Figure 5.24 C well 6	-	Figure 5.25 B well 6	-	Figure 5.26 B well 2	-	Figure 5.27 B well 3
XIV	16-C	N/A		-	Figure 5.24 B well 7	-	Figure 5.24 C well 7	-	Figure 5.25 B well 7	-	Figure 5.26 B well 3	-	Figure 5.27 B well 4
XIV	19-C	-	Figure 5.23 B well 3	+	Figure 5.24 B well 8	-	Figure 5.24 C well 8	-	Figure 5.25 B well 8	-	Figure 5.26 B well 4	-	Figure 5.27 B well 5
XV	2-C	-	Figure 5.23 B well 4	+	Figure 5.24 B well 9	-	Figure 5.24 C well 9	-	Figure 5.25 B well 9	-	Figure 5.26 B well 5	-	Figure 5.27 B well 6
XV	8-C	-	Figure 5.23 B well 5	+	Figure 5.24 B well 10	-	Figure 5.24 C well 10	-	Figure 5.25 B well 10	-	Figure 5.26 B well 6	-	Figure 5.27 B well 7
XVI	2-C	N/A		+	Figure 5.24 B well 11	-	Figure 5.24 C well 11	N/A		N/A		N/A	
XVI	8-C	-	Figure 5.23 B well 6	+	Figure 5.24 B well 12	-	Figure 5.24 C well 12	-	Figure 5.25 B well 11	-	Figure 5.26 B well 7	-	Figure 5.27 B well 8
XVII	6a1-C	N/A		+	Figure 5.24 B well 13	-	Figure 5.24 C well 13	-	Figure 5.25 B well 12	-	Figure 5.26 B well 8	-	Figure 5.27 B well 9
XVII	9a2-C	-	Figure 5.23 B well 7	+	Figure 5.24 B well 14	+	Figure 5.24 C well 14	-	Figure 5.25 B well 13	-	Figure 5.26 B well 9	-	Figure 5.27 B well 10
XVIII	14a2-C	-	Figure 5.23 B well 8	-	Figure 5.24 B well 15	-	Figure 5.24 C well 15	-	Figure 5.25 B well 14	-	Figure 5.26 B well 10	N/A	
XVIII	16a1-C	N/A		+	Figure 5.24 B well 16	+	Figure 5.24 C well 16	-	Figure 5.25 B well 15	-	Figure 5.26 B well 11	-	Figure 5.27 B well 11

3.4.8 Discussion

World-wide movement and commercialisation of live animals can lead to the introduction and establishment of infectious disease in disease-free areas (Lymbery *et al.*, 2014; Peeler *et al.*, 2011). In Europe, appalling examples are the squirrel poxvirus, causing the decline of the native red squirrel and linked to the introduction of the grey squirrel from North America (Rushton *et al.*, 2006), the swim bladder nematode *Anguillicola crassus*, causing the decline of the European eel and linked to the introduction of the Japanese eel (Lymbery *et al.*, 2014), and *Aphanomyces astaci*, causing the decline of European crayfish species and linked to the introduction of North American crayfish which can carry the parasite in the melanised cuticle as chronic infection (Alderman *et al.*, 1984; Unestam, 1972). In particular, for *A. astaci*, from its first introduction in Italy at the end of the 19th century, the parasite rapidly spread across Europe, reaching the Balkan Peninsula, the Black Sea, Turkey and Russia on the east, spreading across the whole of Western Europe, reaching Spain on the south, and entering into the British Isles and Northern Europe (Alderman, 1996; Alderman *et al.*, 1984; Unestam, 1972). Epidemiological studies of crayfish plague dependent on the distinction of *A. astaci* genotypes proved to be a useful tool to better understand the history and spread of this disease (Lilley *et al.*, 1997a; Rezinciuc *et al.*, 2014; Viljamaa-Dirks *et al.*, 2013; Vrålstad *et al.*, 2014) and being able to quickly and reliably identify *A. astaci* genotypes is an important step to the study of crayfish plague. Unfortunately, previously available genotyping techniques present some downsides, including limited sensitivity, reproducibility, and repeatability of the tests, therefore an improved method is needed. In this section, the novel approach of combining bioinformatics and molecular biology techniques, lead to the development of new informative molecular markers to discriminate genotypes of *A. astaci*.

By investigating bioinformatically the genetic diversity of *A. astaci* genotypes (as defined by RAPD-PCR), genotype-specific SNVs within restriction sites were identified and restriction digestion and semi-nested PCR assays were developed in order to exploit this information and successfully identify and discriminate the five known *A. astaci* genotypes from pure cultures and crayfish tissue extracts.

The assays target *A. astaci* nuclear DNA and mitochondrial DNA. The tests targeting nuclear DNA are restriction digestion assays, involving five pairs of primers and the restriction enzyme HhaI. By amplifying the targeted nuclear DNA region containing the selected SNV in the restriction site and digesting the PCR product, the resulting bands pattern on the agarose gel can determine the isolate's genotype.

The tests designed on the mtDNA, instead, present a combination of four restriction digestion assays and a semi-nested PCR. In fact, while for genotypes A, C, D, and E it was possible to detected SNVs within restriction sites, the SNVs unique to genotype B on the mtDNA were not in a restriction site and therefore a semi-nested PCR was developed. The enzymes employed in the restriction digestion were MspI for fragments amplified with the A_MITO primer pair, MseI for fragments amplified with the C_MITO primer pair, AciI for fragments amplified with the D_MITO primer pair and Taq$^{\alpha}$I for fragments amplified with the E_MITO primer pair. For the semi-nested PCR, two primers of equal size were designed, which shared the same nucleotide sequence except for the last base which, for one primer, matches the genotype B SNV, while for the other primer it does not match genotype B SNV. By performing a nested PCR on the PCR product obtained in the first round of PCR, the presence (or absence) of a second band of smaller size in comparison to the original PCR product indicates that the primers matched (or not matched) the variant in the template (Chen and Sullivan, 2003). By using one nested primer at a time in the PCR reaction, the PCR product can be ascribed to genotype B or not.

The assays designed *in silico* worked *in vitro* on pure cultures; however, differences in sensitivity were noted when testing the Cefas historical samples with limited sensitivity for the assays targeting the nuclear DNA.

Tests on gDNA dilutions of *A. astaci* extracted from pure cultures with HhaI primer pairs revealed diverse sensitivity levels, varying from 157 fg for B_HhaI primer pair to 14.4 pg for C_HhaI primer pair (Figure 5.14). These primer pairs were tested on outbreak samples, heavily infected European crayfish, in which the degree of *A. astaci* parasitising the animals is high and this quantity is reflected in the amount of *A. astaci* DNA extracted from the sampled cuticle or

telson. Four out of five primer pairs worked on the outbreaks samples, from which amplification of the targeted regions was achieved and restriction enzymatic assays were performed (section 3.4.6). The differences in the primer pairs performances and the missing amplification of the samples with primer pair E_HhaI (Figure 5.11 B) can be explained to different sensitivity (8.8 pg, Table 5.11), to the presence of contaminants and inhibition factors that influenced the polymerase activity, to sequestered and degraded DNA (Wilson, 1997).

While for the outbreaks samples the amplification of targeted regions was successful with some of the primer pairs designed, for the North American crayfish samples amplification of the targeted regions was not achieved with any primer pairs. In comparison to the outbreak samples, which are from European crayfish, the North American crayfish are less susceptible to developing clinical disease (carriers) and *A. astaci* gDNA extracted from the animal tissues is less abundant. The preferred method to assess the presence of *A. astaci* in carrier tissues is the qPCR, protocol developed by Vrålstad *et al.* (2009) (Tuffs and Oidtmann, 2011). The carrier samples used in this study were positive for *A. astaci* following the qPCR protocol (Dr Birgit Oidtmann, personal communication, March 2016) with an estimated DNA concentration falling between 5 pg and 22 fg (Table 5.1). This amount of gDNA is too small to be detected by the primers designed to target the nuclear DNA.

In contrast, the primer pairs developed to target the mtDNA, revealed a higher sensitivity due to the higher number of mtDNA copies in eukaryotic cells and therefore in the DNA extraction from crayfish tissues (Luo *et al.*, 2011; Waugh, 2007). With these primer sets, successful amplification of the targeted regions from gDNA dilutions of *A. astaci* extracted from pure cultures was obtained at lower template concentration (from 0.39 fg for A_MITO primers pair to 144 fg for C_MITO primers pair), thus increasing the sensitivity of the test by 1000 fold (Figure 5.13; Table 5.11).

As for the nuclear target, the primer pairs targeting the mtDNA were tested on outbreak samples and all five sets of primers amplified the targeted regions, in contrast to the primer pairs targeting the nuclear DNA (*i.e.* E_HhaI primer pair). This could be due to the presence of contaminants and inhibitors or from

degradation of the DNA targeted region in the samples (Wilson, 1997). For the samples that gave a PCR product, restriction digestion and semi-nested PCR assays were performed and genotypes identified (section 3.4.6). Table 3.6 summarises the results of the restriction digestions and the semi-nested PCRs. All outbreaks tested positive for the presence of *A. astaci* genotype B and negative for genotypes C, D, and E, with outbreak PM21018 being positive for *A. astaci* genotype A, alongside genotype B. Sequencing the PCR product of outbreak PM21018 sample number 2 obtained with A_MITO primers confirmed the presence of genotype A in the sample.

The presence of genotype B in the UK has been already reported by experts in the field since the crayfish plague outbreak in 1990 on the River Arrow (Herefordshire) (Alderman, 2003) as revealed by the RAPD-PCR profile of the *A. astaci* pure culture obtained from infected animals and analysed by Lilley *et al.* (1997a), while genotype A has not been detected to date. More recently, the presence of *A. astaci* genotype B in UK has been reconfirmed by microsatellite analysis in the Mochdre Brook (Powys) (James *et al.*, 2017).

A. astaci genotype A is thought to be the first genotype introduced into Europe from a yet unknown North American crayfish back in the 1860s (in Italy) and it has been associated with the first wave of disease in Europe until the 1960s, while the mortalities caused by *A. astaci* after the 1960s are typically associated with genotype B isolates (Alderman, 2003; Huang *et al.*, 1994; Lilley *et al.*, 1997a; Unestam, 1972). Genotype C is based on one single *A. astaci* isolate from *P. leniusculus* native form Canada and imported into Sweden, but was never isolated in European outbreaks (Huang *et al.*, 1994; Jussila *et al.*, 2016); genotype D includes isolates from *P. clarkii* and more recently from *A. pallipes* in Spain (Diéguez-Uribeondo *et al.*, 1995; Rezinciuc *et al.*, 2014); genotype E is the more recent genotype characterised by RAPD-PCR and was isolated from *O. limosus* (Kozubíková *et al.*, 2011).

In the UK, crayfish plague has been linked to the introduction of infected signal crayfish from Sweden and crayfish mortalities associated with the disease were first detected in the summer of 1981 (Alderman, 2003; Alderman *et al.*, 1984; Polglase and Alderman, 1984). The time of the introduction of the parasite and the characterisation of its genotype (B) fit with the historical background of the

disease in UK and can easily explain the reason of the “missing” genotype A in old UK outbreaks.

Another explanation for the lack of genotype A in UK could be due to the difficulty of isolation and maintenance of *A. astaci* pure cultures alive under laboratory conditions (Oidtmann *et al.*, 1999). Moreover, techniques to determine *A. astaci* genotypes from infected animals have been developed only recently (Grandjean *et al.*, 2014; Maguire *et al.*, 2016; Vrålstad *et al.*, 2014) thus the detection of this genotype could have been missed. On the other hand, genotype A has recently been detected by microsatellite analysis in Italy (Pretto *et al.*, 2014) and co-existing with genotype B in Croatia (Maguire *et al.*, 2016). Therefore, the co-existence of these genotypes in the UK cannot be excluded.

Regarding the carrier samples, the primer pairs targeting the mtDNA revealed to be a valid approach to increase the level of detection of *A. astaci* in comparison to the assays targeting the nuclear DNA. Table 3.7 summarises the results of the restriction digests and the semi-nested PCRs. Besides this, from some samples it was not possible to obtain any PCR product and, as explained previously, this could be due to the presence of contaminants, inhibitors, degraded DNA or primer sensitivity (Wilson, 1997). The PCR products obtained were either digested or submitted to semi-nested PCR, depending on the assays, and assigned to genotype B. These findings, beyond confirming the presence of *A. astaci* genotype B in UK water carried by North American crayfish, reflect the original classification and assumption that different *A. astaci* genotypes are carried by different North American crayfish (Grandjean *et al.*, 2014). In this case, the signal crayfish (*P. leniusculus*) sampled would carry genotype B or C isolates (Svoboda *et al.*, 2017). However, recent trial infections of signal crayfish by Aydin *et al.* (2014) revealed that the signal crayfish can be infected and carrying also *A. astaci* genotype A. In the light of the fact that from some of the carrier’s DNA extractions a PCR product was not obtained, it is not possible to exclude the presence of other genotypes in those samples.

A note has to be made about the apparently ambiguous results for the carrier samples III-28C, XVII-9a2C, and XVIII-16a1C (Table 3.7). These samples presented an unusual semi-nested PCR profile and amplification was achieved with both primers (B_MITO_N1S and B_MITO_N2S). These findings, and the

sequenced PCR products, suggests that *A. astaci* genotype B is present but in concomitance with another oomycete or an unknown *A. astaci* genotype, amplified with the semi-nested PCR primer N2S. The presence of unknown genotypes not yet characterised by RAPD-PCR cannot be excluded from these samples, as recent studies suggest the presence of different genotypes in samples tested by microsatellite markers (Viljamaa-Dirks *et al.*, 2014). While the assays developed here can detect low amounts of *A. astaci* gDNA, they were developed based on the typical genotypes defined by RAPD-PCR and cannot confirm the presence of an unknown genotype with only restriction digestions and semi-nested PCRs. Sequencing and analysis of the PCR product obtained with the amplifications is required. On the other hand, the possibility of sequencing the PCR product is an advantage of these assays and can be easily applied to clarify eventual ambiguous results, an approach taken for the carrier samples III-28-C, XVII-9a2-C, and XVIII-16a1-C.

Another note has to be made about the apparent higher sensitivity of some primer pairs when tested on the carrier samples, which, in some cases, presented lower gDNA (quantified by qPCR; Table 5.2) than the sensitivity limits calculated in Table 5.11 (see Figure 5.13 for PCR amplification from *A, astaci* gDNA dilutions). This can be explained considering that these primer pairs do not target specifically only mtDNA from *A. astaci*, but PCR amplification is possible also from mtDNA belonging to other oomycetes (sections 3.4.4 and 3.4.5) that can be present on the crayfish at the moment of the DNA extraction. However, as illustrated by the specificity tests (sections 3.4.4 and 3.4.5), the restriction digestion assays or the semi-nested PCR only work on the selected genotype, or can be easily resolved by sequencing of the PCR product obtained from ambiguous samples.

To date, the only available tool to detect genotypes from crayfish samples was the microsatellite approach. Microsatellite typing bypasses pure cultures, allowing genotyping of *A. astaci* directly from infected host tissues (Grandjean *et al.*, 2014). However, while microsatellites can be a useful tool to detect new undescribed genotypes, they come with some limitations, like the presence of null or partial null alleles due to primers misaligning, the preferred amplification of some alleles rather than others due to the intrinsic competitive nature of the PCR technique or the presence of unknown saprotrophic oomycetes species

can lead to erroneous results (Dakin and Avise, 2004; Grandjean *et al.*, 2014; Jarne and Lagoda, 1996). Moreover, during microsatellite genotyping, differences in alleles loci have been noted by numerous authors in isolates belonging to genotype A, B, and D (James *et al.*, 2017; Maguire *et al.*, 2016; Mrugała *et al.*, 2016; Vrålstad *et al.*, 2014). Another downside of this technique is the agent level required to detect and amplify the DNA from the samples and it is suggested to use this technique when the "agent level", as defined by Vrålstad *et al.* (2009), detected by qPCR is close to or above A4 (corresponding to 24-100 pg) (Grandjean *et al.*, 2014; Maguire *et al.*, 2016; Vrålstad *et al.*, 2014; Vrålstad *et al.*, 2009) while the restriction digestions assays developed here can be used to detect lower concentration of DNA (from 0.39 fg for A_MITO primer pair to 144 fg for C_MITO primer pair on pure cultures Table 5.11, and agent level A2 or below from carrier samples). Moreover, in case of ambiguous results, the PCR products can be purified and sequenced and compared to the genotype-reference sequences, adding a qualitative aspect to the test.

In conclusion, these studies demonstrate the possibility to develop informative molecular markers combining bioinformatics and molecular biology techniques to genotype the causative agent of crayfish plague *A. astaci* and that the options developed here and tested are valuable additions to the genotyping techniques available in the literature, applicable to both pure cultures (nuclear and mitochondrial targets) and crayfish tissues extractions (mitochondrial target). Moreover, limited interference was noted with the oomycetes tested and none with the crayfish gDNA carried in the samples from the gDNA extraction. However, unknown *Aphanomyces* and other oomycetes not tested can be present in the samples, thus attention should be taken when dealing with ambiguous results.

3.5 Case study: an Italian crayfish plague outbreak analysed with assays exploiting single nucleotide variants in the mitochondrial DNA and genotype-specific regions of *Aphanomyces astaci*

In 2012, adult white-clawed crayfish (*Austropotamobius pallipes*) were sampled in northern Italy from wild population and kept under controlled conditions for restocking purpose. From the end of 2013, the presence of an oomycete infection increased the mortality rate, which exceed 80% (Pretto *et al.*, 2014). The oomycete infection was diagnosed to be crayfish plague, caused by *Aphanomyces astaci* genotype A. The presence of *A. astaci* was confirmed by visual inspection of moribund animals for the presence of aseptated hyphae, by isolation of the pathogen on river water-glucose-yeast extract agar and by amplification and sequencing of ITS regions (OIE, 2016a). The level of infection was determined by Pretto *et al.* (2014) following Vrålstad *et al.* (2009) qPCR and the isolate genotype identified by Pretto *et al.* (2014) with RAPD-PCR and microsatellites.

3.5.1 Aims and objectives

This chapter describes application of the genotyping assays developed in this study and described in section 3.4 and 3.3 on samples from this recent well characterised Italian outbreak. The samples were used to assess if the newly developed assays would produce identical results to microsatellite and RAPD-PCR genotyping techniques that were originally utilised to characterise the outbreak, demonstrating the efficacy of the new assays. Moreover, as the agent level in the extracted gDNA from the Italian crayfish tissue samples was higher in comparison to the Cefas samples (Table 5.1 and Table 5.3) the method involving amplification of genotype-specific genomic regions (section 3.3) was tested to explore the potential of this genotyping method from real-world samples.

3.5.2 Materials and Methods

All methods and materials are described in detail in Chapter 2 – Materials and Methods. The samples included gDNA extracts from white clawed crayfish (*A. pallipes*) and from *A. astaci* pure cultures isolated during the outbreak. The isolate was identified by Pretto *et al.* (2014) to belong to genotype A (section 2.1.8). A subset of samples from the Italian outbreak (Table 5.3) were analysed with the genotyping methods described in chapter 3.4 (targeting mtDNA) and section 3.3 (targeting genotype-specific genomic regions), following the PCR protocols described in section 2.3.4.5, section 2.3.4.8 and 2.3.4.9. PCR fragments were checked by gel electrophoresis (section 2.3.1). PCR fragments obtained with the genotype-specific genomic regions primer pair were purified (section 2.3.5), sent to Eurofins MWG Operon commercial sequencing facility to sequence (section 2.3.8) and sequences were visualised and aligned with BioEdit version 7.0.8 (Hall, 1999).

3.5.3 Genotyping assays exploiting SNVs in *Aphanomyces astaci* mtDNA correctly identify the *A. astaci* genotype involved in the outbreak

The primer pairs A, B, C, D, and E_MITO amplified DNA regions from all samples tested (Figure 5.29) and restriction digestion assays were applied on PCR products amplified by A, C, D and E_MITO primer pairs, while semi-nested PCRs were applied to PCR products amplified with B_MITO primer pairs.

The restriction digestions assays for genotype C, D and E were negative (*i.e.* not matching the positive control; Figure 3.36 A 16, B 8 and B 16) indicating the absence of *A. astaci* belonging to these genotypes (Figure 3.36 A 9-15, B 1-7 and B 9-15). Likewise, the semi-nested PCR assay using primers B_MITO_N1S and B_MITO_N2S on the first-round PCR products was negative, indicating the absence of this genotype in the samples (Figure 3.37, lanes 1-7 and 11-17). On the other hand, the restriction digestions of all the PCR products amplified with A_MITO primer pair matched the positive control (*A. astaci* Da – genotype A, Figure 3.36 A, lane 8), confirming the presence of the genotype in the samples (Figure 3.36 A, lanes 1-7).

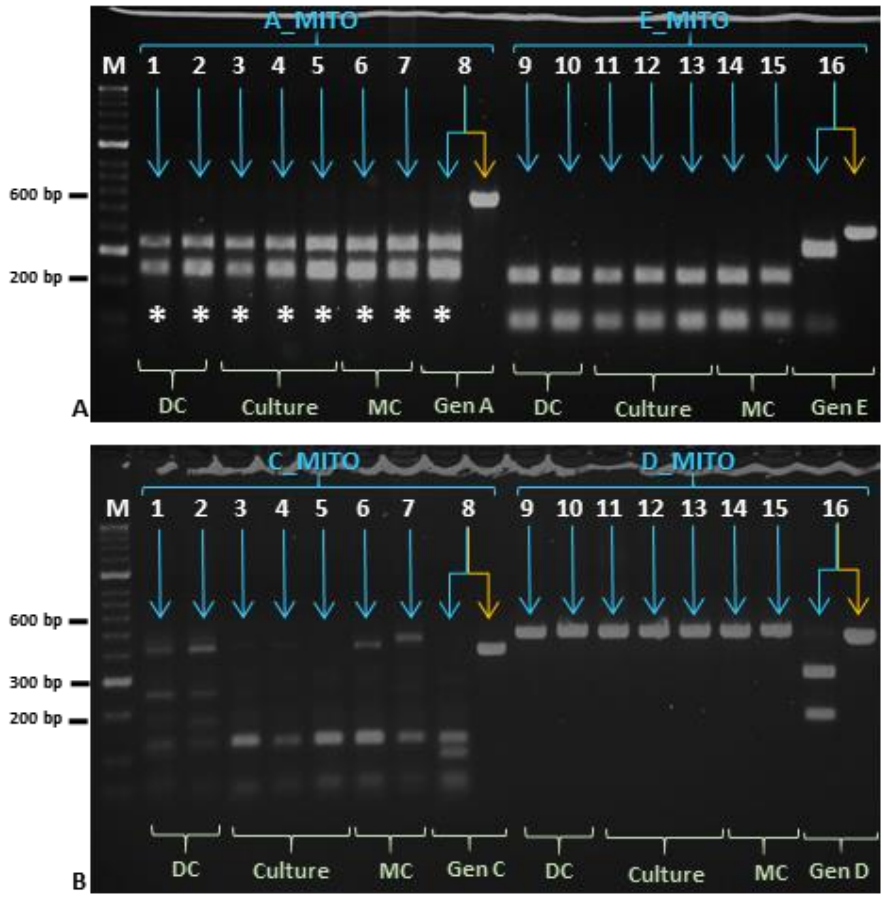


Figure 3.36 Restriction digestions of PCR products from Italian outbreak samples amplified with A, C, D, and E_MITO primer pairs. **A** and **B**: 1-2 and 9-10, gDNA extraction from cuticle of dead *A. pallipes*; 3-5 and 11-13, gDNA extraction from *A. astaci* isolated during the outbreak; 6-7 and 14-15, gDNA extraction from cuticle of moribund *A. pallipes*; **A**8, *A. astaci* Da; **A**16, *A. astaci* KB13; **B**8, *A. astaci* KV; **B**16, *A. astaci* Pc; M, Bioline HyperLadder II. Blue arrows: PCR product after enzymatic digestion. Yellow arrows: original PCR product. *: matching digestion patterns. DC, cuticle from outbreak dead crayfish; Culture, outbreak *A. astaci* pure culture; MC, cuticle from outbreak moribund crayfish.

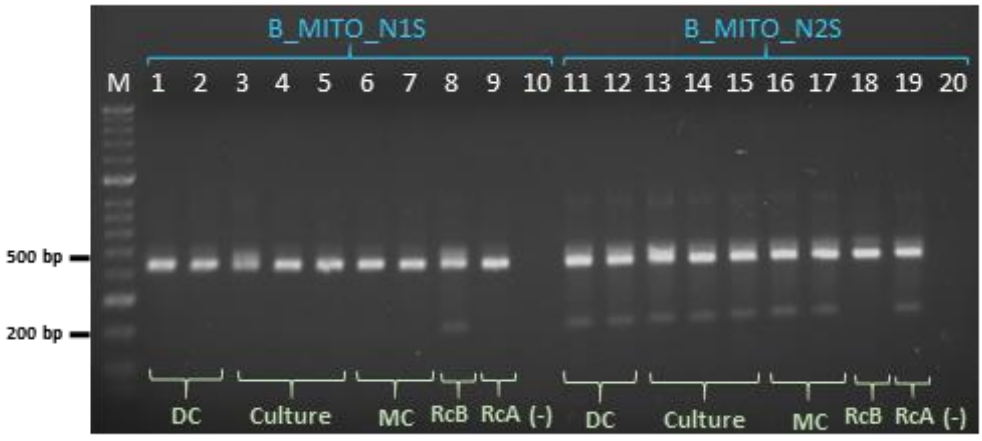


Figure 3.37 Semi-nested PCR on Italian outbreak samples previously amplified with B_MITO primer pairs. M, Bioline HyperLadder II; 1-2 and 11-12, gDNA extraction from cuticle of dead *A. pallipes*; 3-5 and 13-15, gDNA extraction from *A. astaci* isolated during the outbreak; 6-7 and 16-17, gDNA extraction from cuticle of moribund *A. pallipes*; 8 and 18, *A. astaci* D2; 9 and 19, *A. astaci* Da; 10 and 20, negative control. DC, cuticle from outbreak dead crayfish; Culture, outbreak *A. astaci* pure culture; MC, cuticle from outbreak moribund crayfish; RcA, PCR reaction control (genotype A); RcB, PCR reaction control (genotype B); (-), negative control.

3.5.4 Genotyping assays exploiting *Aphanomyces astaci* genotype-specific regions agree with the other genotyping tools

One representative sample (117/ITT/14.α23) of the three gDNA extractions obtained from pure cultures of *A. astaci* isolated during the outbreak was tested with the assays developed in section 3.3 to amplify *A. astaci* genotype-specific regions. Moreover, since the Ct means of the qPCR applied in the samples obtained from infected animals was considerably lower in comparison to the Cefas carriers samples and thus indicating a higher level of agent (*A. astaci*), sample number 17/ITT/14.11 representing gDNA extraction form an infected white clawed crayfish (*A. pallipes*) was included in the analysis.

For both samples, applying PCRs resulted in amplification of the targeted region only with A_unique primer pair. This means that the genotype of *A. astaci* in the samples matches the genotype A specific primer pair, which amplified the targeted region and returned a PCR product. As the genotype of *A. astaci* in the samples did not match genotypes B, C, D and E specific primer pairs, amplification of these genotype-specific regions was not achieved from the samples, but only from the positive controls (Figure 3.38).

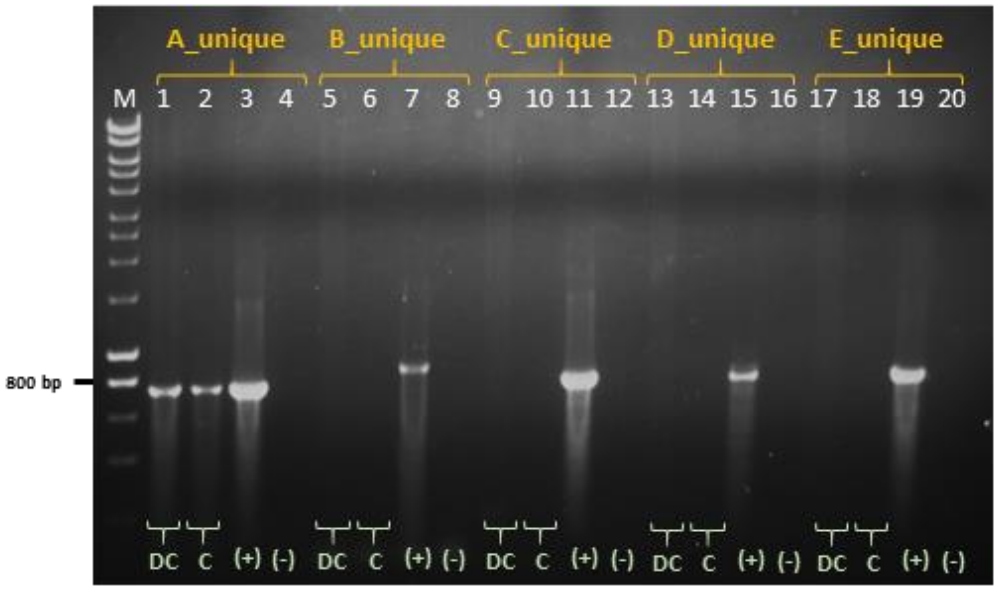


Figure 3.38 Application of genotype-specific genomic regions assays on selected Italian outbreak samples. M, Bioline HyperLadder I; 1, 5, 9, 13 and 17, gDNA extraction from cuticle of dead *A. pallipes* (sample 17/ITT/14.11); 2, 6, 10, 14 and 18, gDNA extraction from *A. astaci* isolated during the outbreak (117/ITT/14.α23); 3, *A. astaci* Da (genotype A); 7, *A. astaci* D2 (genotype B); 11, *A. astaci* KV (genotype C); 15, *A. astaci* Pc (genotype D); 19, *A. astaci* KB13 (genotype E); 4, 8, 12, 16 and 20, negative control.

The PCR products obtained with the A_unique primer pair from the two samples tested were subsequently purified, sequenced and aligned to the genotype A reference sequence to qualitatively characterise the samples. From these alignments (Figure 5.30) the sequences matched the reference genotype A sequence.

3.5.5 Discussion

The performance of the genotyping techniques developed in this study were tested against this well characterised outbreak. Tests on Cefas carriers and outbreak samples with the same approaches were “blind”, as the genotype of *A. astaci* in the samples had not previously been identified. Thus, having the chance to analyse samples from a well characterised *A. astaci* genotype involved in the outbreak was very useful to confirm the potential of the assays. The assays that target SNVs in the mtDNA (restriction digestion assays and semi-nested PCR assay, illustrated in section 3.4) and the assays targeting genotype-specific regions (developed in section 3.3) were tested on a subset of the Italian outbreak samples. In relation to the assays developed to exploit SNVs in the mtDNA, amplification of the targeted regions was achieved from all samples (Figure 5.29). The restriction digestions assays for genotypes C, D and E and semi-nested PCR assay for genotypes B were negative (Figure 3.36 A and B; Figure 3.37), in line with the results from Pretto *et al*. (2014), which identified the Italian outbreak strain as genotype A, based on RAPD-PCR and microsatellites. The restriction digestion assay for genotype A resulted in digested PCR products presenting the same band pattern as the positive control *A. astaci* Da (genotype A), thus correctly identifying the genotype of the outbreak (Figure 3.36 A). The assays developed to amplify genotype-specific genomic regions of *A. astaci* were tested on a subset of samples and amplification of a genomic region was obtained for both samples only with the primer pair specific for genotype A (Figure 3.38). This result was also confirmed by analysis of the sequenced PCR products, which matched the *A. astaci* genotype A reference sequence (Figure 5.30).

In section 3.3, the application of the assays to amplify genotype-specific genomic regions on samples derived from historical crayfish tissues was not

successful. As the nuclear DNA regions targeted by these primer pairs are considerably longer in comparison to the regions on the mtDNA, making the amplification more difficult, especially from samples that have been stored for a long time and in which the *A. astaci* DNA might have been degraded due to storing conditions or enzymatic degradation (Wilson, 1997). Unfortunately, the concentration of *A. astaci* gDNA in the Cefas outbreaks samples (Table 5.2) was not determined by qPCR, therefore a direct comparison of the results is not possible.

The work described in this chapter confirms that the restriction digestion assays and the semi-nested PCR assay work on gDNA extracted from infected crayfish and pure cultures not previously included in the development of the assays. Moreover, the amplification of genotype-specific regions can indeed be applied to gDNA extracted from crayfish tissues and can be successfully used to identify the genotype of the *A. astaci* involved in the outbreak. Being able to subsequently sequence the PCR products obtained using these approaches adds qualitative analysis of the sequences amplified by the PCR and is an additional value to whole assays, in comparison to the RAPD-PCR methods whose bands cannot directly be sequenced.

3.6 Insights from whole-genome sequencing

Whole genome sequencing (WGS) has been successfully applied to various fields of biology in order to better understand pathologies, host pathogen interactions, microbial evolution, epidemiology, characterisation of outbreaks and populations genetics (Grad *et al.*, 2012; Gudbjartsson *et al.*, 2015; Hao *et al.*, 2012; Jiang *et al.*, 2013; Liti *et al.*, 2009; Sabat *et al.*, 2013; Studholme *et al.*, 2011). WGS has been also applied to identify differences between closely related organisms and to develop genotyping techniques (Boutet *et al.*, 2016; Kim *et al.*, 2016; Pendleton *et al.*, 2013; Roetzer *et al.*, 2013). This concept has been followed in this study in order to develop new genotyping assays able to distinguish all known *A. astaci* genotypes. The assays can detect the genotypes directly from infected animals (especially carriers) bypassing the pure culture step, adding valuable tools/information to the genotyping methods already available in the field.

3.6.1 Aims and objectives

The aims of the data presented below were to compare the assembled genomes of *A. astaci* and the other *Aphanomyces* species sequenced in this study in order to detect genomic differences among the *A. astaci* RAPD-based genotypes and between species (investigated using dnadiff), and the differences in the internal transcribed spacer and mitochondrial DNA (mtDNA) copies elucidating the different sensitivity of the genotyping assays developed in the study (section 3.3 and 3.4).

3.6.2 Materials and methods

All methods and materials are described in detail in Chapter 2 – Materials and Methods. Briefly, gDNA was extracted from 15 isolates (Table 2.4) held in the OCC (Table 2.1), submitted for WGS and sequenced using the ultra-high-throughput Illumina HiSeq 2500 System to generate paired 100, 150 or 300 bp reads (section 2.3.2 and 2.3.3). Illumina reads were quality checked and trimmed (section 2.4.3). In order to choose the best assembler for the datasets generated, WGS data from a subset of sequenced isolates was used to

undertake an initial evaluation of 3 assemblers (section 2.4.4) and best assembler chosen by best contiguity and completeness, assessed with QUAST (Gurevich *et al.*, 2013). and BUSCO (Simão *et al.*, 2015) (section 2.4.5). In order to identify genotype differences, dnadiff (Kurtz *et al.*, 2004) was used to compare *A. astaci* and other *Aphanomyces* genomes assembled in this study, *A. astaci* APO3 and *A. invadans* NJM9701 sequenced and assembled by the Broad Institute (GenBank assembly accessions: GCA_000520075.1 and GCA_000520115.1) (2.4.12).

The estimated ITS copy number in the genome of *A. astaci* D2 was calculated by comparing the depth of coverage of the ITS region in the genome to the mean depth of coverage of the whole assembly. The mean coverage of the whole assembly was calculated by (number of reads) x (reads insert length) / (size of genome). To compare the mean depth of coverage of the mitochondrial DNA against the coverage of the whole assembly, the mean coverage of *A. astaci* D2 mitochondrial genome was calculated by (number of reads aligned to the whole mitochondrial assembly) x (reads insert length) / (size of mitochondrial genome).

3.6.3 Results on tested assemblers

A. astaci 457 and 197901, *A. frigidophilus* AP5 and RP2, *A. invadans* NJM0002, and *A. invadans*-like NJM9510 paired-end reads were assembled using three different genome assemblers (SOAPdenovo, Velvet and SPAdes) to test their performance on the datasets produced in this study and assessed for contiguity and completeness with QUAST and BUSCO.

For all the six set of paired-end reads, the assemblies with better contiguity (N_{50} length, average length of contigs covering 50% of the genome; N_{50} number, number of contigs covering 50% of the genome) values were the ones returned by SPAdes, data agreeing with BUSCO analysis, which detected higher number of complete single-copy orthologs in the assemblies generated by SPAdes. The only exception was *A. invadans*-like NJM9510, whose BUSCO values were higher for the assembly generated by Velvet. Overall, the worst contiguity and BUSCO values were for the assemblies generated by SOAPdenovo. Table 3.8

summarises contiguity and completeness values for the assemblies produced by the three assemblers.

SPAdes assembler was therefore chosen to assemble all the paired-end reads generated from the remaining *A. astaci* isolates as the assemblies produced by SPAdes presented better contiguity and completeness in comparison to the assemblies produced by Velvet and SOAPdenovo.

Table 3.9 summarises contiguity and completeness values calculated for the remaining *A. astaci* isolates paired-end reads assembled with SPAdes. These assemblies were compared to the assessments made on the publicly available reference assemblies of *A. astaci* APO3 genotype D (GenBank assembly accession number: GCA_000520075.1) and *A. invadans* NJM9701 (GenBank assembly accession number: GCA_000520115.1) assembled by the Broad institute and retrieved from NCBI website.

Table 3.8 Assemblers assessment by contiguity and completeness values. *A. astaci* 457 and 197901, *A. frigidophilus* AP5 and RP2, *A. invadans* NJM0002 and *A. invadans*-like NJM9510 paired-end reads were assembled with SPAdes, Velvet and SOAPdenovo genome assemblers. Contiguity and completeness values were calculated with QUAST on contigs. N_{50} length: average length of contigs covering 50% of the genome. N_{50} number: number of contigs covering 50% of the genome. Total length: total length in base pair of the contigs. Universal Single-Copy Orthologs (USCOs) are expressed in % of orthologs found on a total of 429 searched.

Isolate	**Assembler**		
***A. invadans* NJM0002**	**SPAdes**	**Velvet**	**SOAPdenovo**
N_{50} length	14191	7964	1211
N_{50} number	790	1175	6310
Total length (bp)	51805647	46763934	39334030
Number of contigs	11218	14623	33295
Complete Single-Copy USCOs	73%	69%	24%
Complete Duplicated USCOs	9.3%	9.7%	2.7%
Fragmented USCOs	8.8%	12%	31%
Missing USCOs	17%	17%	44%
***A. astaci* 197901**	**SPAdes**	**Velvet**	**SOAPdenovo**
N_{50} length	6184	1617	1450
N_{50} number	2941	10712	10079
Total length (bp)	70928142	62417636	55991194
Number of contigs	22957	45590	44652
Complete Single-Copy USCOs	67%	64%	37%
Complete Duplicated USCOs	9.5%	8.8%	5.8%
Fragmented USCOs	13%	16%	28%
Missing USCOs	18%	19%	34%
***A. astaci* 457**	**SPAdes**	**Velvet**	**SOAPdenovo**
N_{50} length	4804	1751	1545
N_{50} number	3710	9003	8522
Total length (bp)	70068905	59146102	52903652
Number of contigs	26929	40718	40458
Complete Single-Copy USCOs	59%	55%	36%
Complete Duplicated USCOs	8.8%	9%	6.9%
Fragmented USCOs	21%	23%	31%
Missing USCOs	19%	20%	31%

Isolate	**Assembler**		
***A. frigidophilus* AP5**	**SPAdes**	**Velvet**	**SOAPdenovo**
N_{50} length	8408	3650	2720
N_{50} number	2446	5822	6231
Total length (bp)	77438100	75572111	64923921
Number of contigs	19384	30916	34598
Complete Single-Copy USCOs	71%	67%	49%
Complete Duplicated USCOs	29%	35%	18%
Fragmented USCOs	11%	15%	24%
Missing USCOs	16%	16%	25%
***A. frigidophilus* RP2**	**SPAdes**	**Velvet**	**SOAPdenovo**
N_{50} length	6994	2763	1982
N_{50} number	3775	10642	11348
Total length (bp)	93756549	101718951	82304015
Number of contigs	22603	51647	54046
Complete Single-Copy USCOs	67%	62%	36%
Complete Duplicated USCOs	32%	37%	16%
Fragmented USCOs	14%	17%	30%
Missing USCOs	17%	19%	32%
***A. invadans*-like NJM9510**	**SPAdes**	**Velvet**	**SOAPdenovo**
N_{50} length	3902	2151	1347
N_{50} number	1811	3463	4592
Total length (bp)	46398419	38811366	32340985
Number of contigs	21553	22443	25857
Complete Single-Copy USCOs	42%	49%	14%
Complete Duplicated USCOs	4.6%	6.2%	1.6%
Fragmented USCOs	16%	17%	13%
Missing USCOs	41%	33%	71%

Table 3.9 Contiguity and completeness values of the remaining *A. astaci* isolates and the reference assemblies (*A. astaci* APO3 - GenBank assembly accession number: GCA_000520075.1; *A. invadans* NJM9701 - GenBank assembly accession number: GCA_000520115.1). Contiguity and completeness values were calculated with QUAST on contigs. N_{50} length: average length of contigs covering 50% of the genome. N_{50} number: number of contigs covering 50% of the genome. Total length: total length in base pair of the contigs. Universal Single-Copy Orthologs (USCOs) are expressed in % of orthologs found on a total of 429 searched.

Isolate (genotype)	N_{50} length	N_{50} number	Total length (bp)	Number of contigs	Complete Single-Copy USCOs	Complete Duplicated USCOs	Fragmented USCOs	Missing USCOs
***A. astaci* Da (A)**	14574	921	51803019	7947	72%	10%	10%	16%
***A. astaci* SV (A)**	13249	1037	51843548	8531	67%	8.3%	14%	17%
***A. astaci* D2 (B)**	8846	1605	53438091	12012	69%	9.5%	14%	16%
***A. astaci* SA (B)**	8040	1732	53229505	12595	68%	8.8%	14%	16%
***A. astaci* Si (B)**	8469	1706	53510754	12376	69%	9%	14%	15%
***A. astaci* YX (B)**	7520	1915	53610083	13242	72%	9.5%	11%	15%
***A. astaci* KV (C)**	8740	1626	53686410	11919	69%	9.5%	13%	16%
***A. astaci* Pc (D)**	16126	868	52812792	7831	73%	10%	9.5%	16%
***A. astaci* KB13 (E)**	2121	12164	104586921	66018	52%	10%	27%	19%
***A. astaci* APO3 (D)**	36439	421	58567734	4648	77%	10%	6.9%	15%
***A. invadans* NJM9701**	30557	335	41381033	5193	76%	8.1%	7.4%,	15%

3.6.4 Identification of differences between *A. astaci* genotypes and between *Aphanomyces* species with the aid of dnadiff

To assess the similarity and differences between *A. astaci* genotypes, dnadiff was used to compare the assembled genomes of the 11 *A. astaci* isolates sequenced and assembled in this study and *A. astaci* APO3 sequenced and assembled by the Broad Institute (GenBank assembly accession: GCA_000520075.1). dnadiff compares highly similar genomes (*e.g.* comparing isolates belonging to the same species) and returns a report which includes alignments statistics (*e.g.* % of aligned/unaligned sequences and % of sequence identities), used to understand similarity and differences between the assemblies.

Moreover, *A. astaci* genome assemblies and one representative assembly for each *A. frigidophilus*, *A. invadans* and *A. invadans*-like species were compared to *Aphanomyces* assemblies generated in the present study in order to identify the minimum level of genome similarity between two isolates that could indicate appartenence to a determined species (Table 5.14, Figure 3.39, Figure 3.40, and Figure 5.31).

Regarding the analysis on the % of sequence identities between *A. astaci* genotypes (Table 5.14, Figure 3.39 and Figure 5.31), all isolates belonging to a specific genotype had high percentage (~99 %) of matching sequences. Unfortunately, only one isolate each were available for genotype C and E and comparison with other isolates belonging to the same genotypes was not possible.

Regarding the analysis on the % of sequence identities between *Aphanomyces* species (Table 5.14, Figure 3.39, Figure 3.40, and Figure 5.31), all isolates belonging to a species presented high level of similarity of the assemblies, which was above 95%. In particular, *A. invadans* isolates assemblies matched for 99%, while *A. frigidophilus* isolates and *A. astaci* isolates matched for 95%. Unfortunately, only one isolate of *A. invadans*-like was available and comparison with other isolates belonging to this species was not possible. However, while the genome assembly of this species is 88 and 87% similar to *A. astaci* and *A. invadans* respectively, it's 91% similar to *A. frigidophilus*.

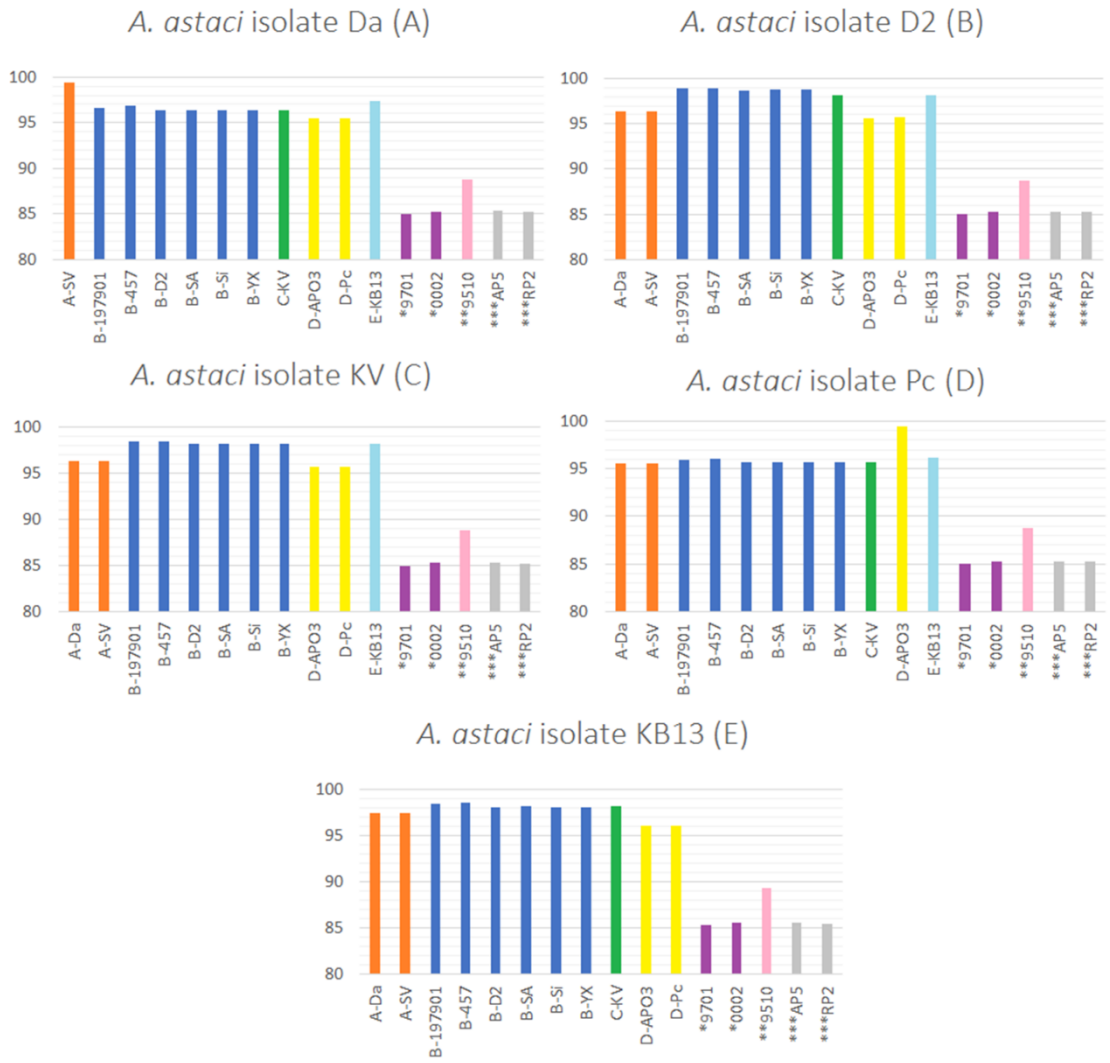


Figure 3.39 Dnadiff percentage of sequence identity of a representative isolate for each genotype and other *Aphanomyces* species. Y axis: % of sequence identity; X axis: isolates; orange bars: *A. astaci* SV and Da (genotype A); dark blue bars: *A. astaci* D2, Si, SA, YX, 457, 197901 (genotype B); green bars: *A. astaci* KV (genotype C); yellow bars: *A. astaci* APO3 and Pc (genotype D); light blue bars: *A. astaci* KB13 (genotype E); purple bars: *A. invadans* NJM9701 and NJM0002; pink bars: *A. invadans*-like NJM9510; grey bars: *A. frigidophilus* AP5 and RP2. *, *A. invadans*; **, *A. invadans*-like; ***, *A. frigidophilus*.

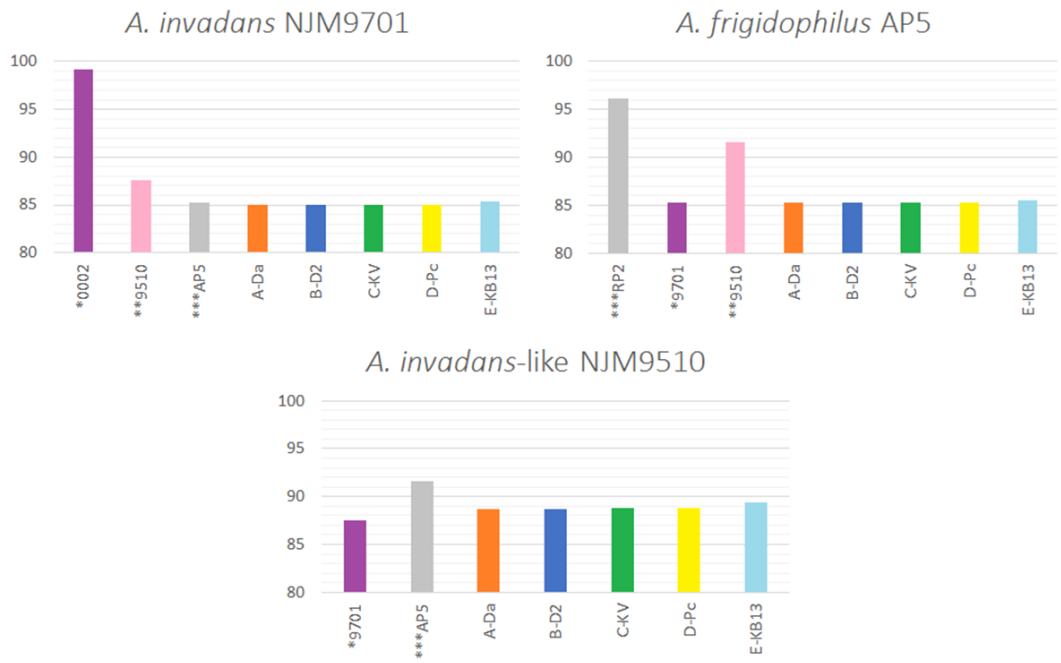


Figure 3.40 Dnadiff percentage of sequence identity between a representative isolate of *Aphanomyces invadans, A. frigidophilus,* and *A. invadans*-like species and representative assemblies generated in the present study. Y axis: % of sequence identity; X axis: isolates; orange bars: *A. astaci* Da (genotype A); dark blue bars: *A. astaci* D2 (genotype B); green bars: *A. astaci* KV (genotype C); yellow bars: *A. astaci* Pc (genotype D); light blue bars: *A. astaci* KB13 (genotype E); purple bars: *A. invadans* NJM9701 and NJM0002; pink bars: *A. invadans*-like NJM9510; grey bars: *A. frigidophilus* AP5 and RP2. *, *A. invadans*; **, *A. invadans*-like; ***, *A. frigidophilus*.

3.6.5 Infering the average coverage of *A. astaci* D2 whole and mitochondrial genomes and the ITS copy number

To understand the potentiality of targeting the mitochondrial genome to develop genotyping tools instead of targeting the nuclear genome, the average depth of coverage for the whole genome assembly of *A. astaci* D2 was calculated and compared to the average depth of coverage of *A. astaci* D2 mitochondrial genome assembly (Table 3.10) and to the number of ITS regions expected in a genome assembly.

Table 3.10 Average coverage of A. *astaci* D2 whole and mitochondrial genome assemblies.

	Size (bp)	**Reads**	**Average depth of coverage**
Mitochondrial genome	33358	1035780	3105
Whole genome	53438091	26157910	49

The ITS region coverage for *A. astaci* D2 genome assembly was equal to 3171 reads, value comparable to the mitochondrial coverage. By dividing this value by the average depth of coverage of whole genome, it is possible to obtain the representation of this region in the genome, that in this case is equal to 65.

3.6.6 Discussion

Whole genome sequencing (WGS) has enabled the scientific community to build, study and decipher the genomes of a multitude of different organisms, including viruses (Ackermann and Kropinski, 2007), prokaryotes *i.e.* bacteria and archaebacteria (Anderson *et al.*, 2009; Ivanova *et al.*, 2003), unicellular eukaryotes *i.e.* yeasts and heterokonts (Denoeud *et al.*, 2011; Jiang *et al.*, 2013; Martinez *et al.*, 2004) and multicellular eukaryotes *i.e.* vertebrates, fungi, plants and invertebrates (Birol *et al.*, 2013; IHGSC, 2004; Myers *et al.*, 2000; Stajich *et al.*, 2010; Wang *et al.*, 2011). WGS allows and facilitates the study of diseases and microbial evolution, with an important role also in epidemiology and characterisation of disease outbreak (Grad *et al.*, 2012; Gudbjartsson *et al.*, 2015; Hao *et al.*, 2012; Jiang *et al.*, 2013; Liti *et al.*, 2009; Metzker, 2010; Sabat *et al.*, 2013; Studholme *et al.*, 2011). WGS has been applied to identify differences between closely related organisms and to develop genotyping techniques (Boutet *et al.*, 2016; Kim *et al.*, 2016; Pendleton *et al.*, 2013; Roetzer *et al.*, 2013), a concept followed in this study to develop new genotyping assays able to distinguish all known *A. astaci* genotypes. Therefore, all *A. astaci* isolates available in Cefas OCC (Table 2.1) alongside three other *Aphanomyces* species (*A. invadans, A. frigidophilus* and *A. invadans*-like) were sequenced. The WGS data obtained from a subset of sequenced isolates were then assembled with three different assemblers (SPAdes, SOAPdenovo and Velvet) and the performance of the assemblers assessed by determining the completeness and the contiguity of the assemblies. This is an important step to assess the quality of a newly assembled genome, especially if a reference genome is not available, as the assembling process is not perfect and error-free (Parra *et al.*, 2009; Paszkiewicz and Studholme, 2010; Salzberg and Yorke, 2005; Studholme, 2016). The most used statistic parameters to evaluate the contiguity of an assembly are the contigs (or the scaffolds) N_{50} number and the N_{50} length (Paszkiewicz and Studholme, 2010). N_{50} number represents the number of contigs (or scaffolds) needed to cover 50 % of the genome. The lower this value is, the longer are the contigs (or scaffolds) that construct that assembly; the N_{50} length is the average length of contigs (or scaffolds) needed to complete 50 % of the genome, thus the higher this value is, the longer are the contigs (or scaffolds) that form the assembly. These metrics are a good

indication of contiguity but may not reflect the completeness and the accuracy of the assembly (Baker, 2012; Paszkiewicz and Studholme, 2010; Salzberg and Yorke, 2005) and for this reason the assemblies were analysed with BUSCO, software that estimates the completeness of an assembled genome by performing a census of conserved single-copy genes that ought to be found in the genomes of particular lineages of organisms. By mapping these genes to a genome, BUSCO assesses their presence in single or multi-copy and their completeness, giving an indication of how well the genome has been assembled (Simão *et al.*, 2015). For five of six set of paired-end reads tested (*A. astaci* 457 and 197901, *A. frigidophilus* AP5 and RP2, and *A. invadans* NJM0002) SPAdes generated better assemblies with higher contiguity, reflected by BUSCO's analysis likewise, which detected higher number of complete single-copy orthologs. The only exception was *A. invadans*-like NJM9510, whose BUSCO values were higher for the assembly generated by Velvet. Overall, the worst contiguity and BUSCO values were for the assemblies generated by SOAPdenovo (Table 3.8). SPAdes assembler was therefore chosen to assemble all the paired-end reads generated from the remaining *A. astaci* isolates.

The assemblies (size, contiguity, and completeness) were then compared to *A. astaci* and *A. invadans* reference genomes currently available on NCBI database. Overall, the size of *A. astaci* isolates Da, SV, D2, SA, Si, YX, KV, and Pc assemblies is smaller (51-53 Mbp) but comparable to the *A. astaci* APO3 reference assembly (58 Mbp) as are the % of duplicated and missing USCOs. The % of complete single-copy USCOs is instead slightly lower (67-73 % vs 77 %) while the % fragmented USCOs is slightly higher (9.5-14 % vs 6.9 %) in comparison to *A. astaci* APO3. These findings possibly indicate that the *A. astaci* isolates Da, SV, D2, SA, Si, YX, KV, and Pc assemblies are more fragmented but overall similar to *A. astaci* APO3. *A. astaci* 457, 197901 and KB13 presented instead bigger assemblies (70 Mbp and 104 Mbp respectively). Considering BUSCO's results, KB13 and 457 assemblies presented a higher number of fragmented USCOs, a similar number of duplicated USCOs and a slightly higher number of missing USCOs in comparison to APO3 reference assembly, while 197901 assembly presented a higher number of duplicated USCOs. This analysis could indicate a loss of information (due to regions of the

genome that have not been sequenced by the Illumina platform or that have been discarded by the assembler) and the presence of misassembled or duplicated regions that the assembler SPAdes has not been able to resolve. In these last cases, while high heterozygous sequences typical of diploid organisms create different paths in the de Bruijn graph (used by SPAdes assembler) and are inaccurately resolved as multiple sequences creating a longer assembly, long repetitive reads are seen as identical by the assembler which crunches and collapses them causing misassembles (Kajitani *et al.*, 2014; Phillippy *et al.*, 2008; Salzberg and Yorke, 2005). The differences noted between assemblies built in this study and *A. astaci* APO3 genome assembled by the Broad Institute, could originate from the different read lengths of the paired-end reads from the Illumina platform. While paired-end reads of 100-150 bp were generated in the present study, the ones generated for the APO3 assembly were longer, between 3 to 5 kbp (Makkonen *et al.*, 2016), and longer reads can resolve long repeats improving the overall assembly (Ribeiro *et al.*, 2012). Generating a correct and complete assembly is important especially for downstream analysis of the organism's genome (*e.g.* gene prediction and annotation) and the assembly can be improved for example also by combining more libraries generated with different technologies (*e.g.* PacBio) (Paszkiewicz and Studholme, 2010). Having a whole comprehensive and perfect assembly was beyond the objectives of this study and the datasets generated were good enough to correctly identify the genotype-specific SNVs (sections 3.1 and 3.4.5) and genotype-specific genomic regions (section 3.3) used in the study to develop *A. astaci* new genotyping tools.

The *A. astaci* genomes assembled in this study, were furthemore analysed with dnadiff in order to identify any relevant differences between genotypes and between isolates that could lead to the development of new genotyping techniques (leading to the design of the genotyping tool described in section 3.3). At present, five genotypes of *A. astaci* have been isolated and genetically characterised by RAPD-PCR, AFLP and microsatellites analysis: genotype A has been connected to the first introduction of crayfish plague in Europe; genotype B has been identified after the introduction of *P. leniusculus* in Europe; genotype C comprises only one isolate from *P. leniusculus* native form Canada and imported into Sweden; genotype D includes mainly isolates from *P.*

clarkii; genotype E was firstly isolated from *O. limosus* (Diéguez-Uribeondo *et al.*, 1995; Grandjean *et al.*, 2014; Huang *et al.*, 1994; Kozubíková, *et al.*, 2011; Rezinciuc *et al.*, 2014). From the dnadiff analysis, isolates Da and SV, both belonging to genotype A, presented a high % of similarity (99.41 %). While isolate SV was originally from Sweden (Table 2.1), it is not possible to trace the origin of isolate Da and compare the isolates' histories. Isolates belonging to genotype B (D2, Si, SA, YX, 457, and 197901) presented a high range of sequence similarity, ranging from 98.72% to 99.3%. Differently to isolates D2, SA, and YX, isolates 457 and 197901 showed higher similarity between each other. Both were isolated in the UK (Alderman, 1996; Lilley *et al.*, 1997a; Tuffs and Oidtmann, 2011), which could explain the high similarity. As for genotype A and B, isolates belonging to genotype D (Pc and APO3) presented a high range of sequence similarity (99.47 %). Genotype B, C, and E isolates share a high percentage of similar sequences with each other in comparison to genotype A and D, agreeing with the SNVs analysis conducted in section 3.1. These findings confirm once again that the genetic differences separating the 5 genotypes could reflect their history and their host-parasite interaction as illustrated in section 3.1. From the dnadiff analysis, another feature can be observed (especially in the isolates belonging to genotype B). Isolates belonging to the same genotype, even if sharing high similarity, show some variability that could harbour important features or information regarding the single isolate identity. For example, infection studies on genotypes A and B indicate differences in virulence between genotypes and, more importantly, between isolates belonging to the same genotype (Becking *et al.*, 2015; Makkonen *et al.*, 2014; Makkonen *et al.*, 2012; Viljamaa-Dirks *et al.*, 2016). Moreover, during microsatellite genotyping, differences in alleles loci have been noted by numerous authors in isolates belonging to genotype A, B, and D (James *et al.*, 2017; Maguire *et al.*, 2016; Mrugała *et al.*, 2016; Vrålstad *et al.*, 2014). These differences, together with the variabilities noted in this study in the dnadiff comparisons, could harbour the potential presence of closely related sub-genotypes or isolate-specific characteristics that could improve the degree of resolution in epidemiology studies.

For each *A. frigidophilus*, *A. invadans*, *A. invadans*-like species and one representative *A. astaci* genome assembly of each genotype were compared to

the *Aphanomyces* assemblies generated in the present study to identify the minimum level of genome similarity between two isolates needed to ascribe an unknown isolate to a species (3.6.4). From this analysis, all isolates belonging to a species presented high level of assembly similarity, which was above 95 % (Table 5.14, Figure 3.39, Figure 3.40, and Figure 5.31). Comparing members of the same species was possible only for *A. astaci, A. invadans*, and *A. frigidophilus* as more than two isolates were available for WGS. *A. invadans* isolates assemblies matched for 99 %, while *A. frigidophilus* isolates and *A. astaci* isolates matched for 95 %. The similarity between *A. invadans* genome assemblies is as high as the similarities between *A. astaci* isolates belonging to the same genotype (between 98 to 99 %). This finding reinforces the conclusions of Lilley *et al*. (1997b), who investigated the genetic diversity of 20 *A. invadans* isolates by applying RAPD-PCR with 14 primers on gDNA and restriction fragment length polymorphisms on amplified ITS regions from the same pure cultures. With these two genotyping methods, the 20 isolates did not show enough genomic variation to identify different genotypes, suggesting the presence of a single *A. invadans* clone that have been distributed world-wide by commercial trades of aquatic animals. However, as for *A. astaci* isolates belonging to the same genotype, *A. invadans* isolates NJM9701 and NJM0002 genome assemblies compared in the current study presented a degree of low variability that could indicate the presence of isolate-specific genomic features, which could not be identified with the genotyping techniques previously applied. Interestingly, *A. frigidophilus* genome assemblies showed only 96 % of sequence similarity, result comparable to the sequence similarity between *A. astaci* isolates belonging to difference genotypes (3.6.4). This genomic variation could indicate the presence of genotypes in *A. frigidophilus* species, phenomenon not yet investigated for this species. Unfortunately, only one isolate of *A. invadans*-like was available for WGS and comparison with other isolates belonging to this species was not possible. However, while the genome assembly of this species presents 87-88 % of sequence similarity to *A. invadans* and *A. astaci* respectively, it is 91 % similar to *A. frigidophilus*. The phylogenetic analyses based on the internal transcribed spacers of ribosomal DNA (ITS) (3.2.3), the cytochrome oxidase subunit I (COI) (3.2.5), and the D1-D2 regions of the large subunit ribosomal DNA (LSU) (3.2.4) show that *A. invadans*-like is closer related to the opportunistic/saprotrophic *Aphanomyces*

species rather than to the animal parasitic ones (*i.e. A. astaci* and *A. invadans*). *A. frigidophilus* isolates have been collected from a wide range of substrates, including salmonid eggs (Kitancharoen and Hatai, 1997) and dead freshwater crayfish (Ballesteros *et al*., 2006), but infection studies in crayfish failed to demonstrate its pathogenicity to this species. It is still uncertain whether this species is parasitic or opportunistic (Diéguez-Uribeondo *et al.*, 2009) even if in ITS based phylogenetical analysis of the *Aphanomyces* genus *A. frigidophilus* groups within the animal parasites species (Figure 1.2; sections 3.2.3, 3.2.4, and 3.2.5). As both *A. frigidophilus* and *A. invadans*-like pathogenicity against a vertebrate (*e.g.* fish) or invertebrate (*e.g.* crayfish) have not been fully defined, they could indeed be opportunistic/saprotrophic *Aphanomyces* species and therefore share more genome sequence similarity than with animal parasitic *Aphanomyces* species.

The sequences difference noted between isolates belonging to the same species (*A. astaci* and *A. frigidophilus* isolates) raised the question if 95 % of genome similarity would be enough to include one or more individuals in the same species. For example, humans and chimpanzees share about 96 % of DNA (The Chimpanzee Sequencing and Analysis Consortium, 2005; Varki and Altheide, 2005), while humans and bonobos share about 98 % of DNA (Püfer *et al*., 2012), but they are distinct species. Could this mean, for instance, that *A. astaci* genotypes possess enough genomic variation to be distinguished in species rather than genotypes? In biology, there are many definitions of “species” with the widely accepted concepts that a species is a group of individuals presenting similar phenotype and DNA which can distinguish them from another group of individuals and which can exchange genes by sexual reproduction within the same group but not between groups (Birky *et al*, 2010). Species are the fundamental units of biology, yet defining the notion of species is one of the most argued concept in biology. Some definitions of species include individuals that can interbreed, producing viable and fertile offspring (Mayr, 1942; Wright, 1940), individuals forming phenetic groups, which present qualitative and quantitative differences from other individuals (Michener, 1970; Sneath and Sokal, 1973), and individuals presenting alleles inherited by a common ancestral parent and not shared with other individuals (Baum and Shaw, 1995). NGS, WGS, and sequence analysis now allow a deeper

comparison between individuals belonging to closely related species or between strains and isolates belonging to the same species. The use of these techniques in species identification, established other criteria to define a species, for example by counting and comparing the SNVs in the whole genome of two individuals, by identifying genome regions with a high divergence level between species, or by annotating and comparing important genome features, like functional elements (Frazer *et al.*, 2003; Turner *et al.*, 2005; Harrison and Larson, 2014). However, different genomic regions evolve at different speed and it can be challenging to choose universal thresholds of sequence length and percentage of identity that can distinguish individuals into species in all cases (Frazer *et al.*, 2003). Varieties of organisms can raise from ancestral parents and become species by gradually accumulating changes (*e.g.* phenotypic or genetic), but the amount of differences (genomic or phenotypic) needed to consider two individuals belonging to different species cannot be defined universally (Birky *et al*, 2010). Until recently, to attribute a sudden crayfish mortality to *A. astaci*, the species of the animal pathogenic *Aphanomyces* involved in the outbreak had to be investigated by isolation of the pathogen and transmission/infection trials, as there are no morphological features able to distinguish *A. astaci* from other saprotrophic/opportunistic *Aphanomyces* species (OIE, 2016a; Vrålstad *et al.*, 2014). The development of PCR and qPCR (Oidtmann *et al.*, 2006; Vrålstad *et al.*, 2009) targeting *A. astaci* ITS regions have improved the crayfish plague diagnostic tool set, allowing reliable species identification of *A. astaci* pure cultures, and fresh and preserved crayfish samples infected with *A. astaci*. As previously shown in section 3.2, the five known *A. astaci* genotypes cannot be distinguished by the analysis of the ITS, LSU, and COI, indicating that the isolates form one species (*A. astaci*), even if inconsistent intraspecific nucleotide variants were observed in all three molecular markers. However, COI sequences obtained from isolates belonging to genotype D presented consistent intraspecies variants and could separate genotype D from the other genotype (Figure 3.13). Moreover, *A. astaci* isolates belonging to genotype D are adapted to higher temperatures, in contrast to the other four known genotypes which are adapted to lower temperatures (Rezinciuc *et al.*, 2014). The occupation of different environmental niches, the adaptation to different environmental conditions, and the physical separation between environments, are few of the speciation factors allowing the

independent evolution of organisms into species (Barraclough *et al*., 2003; Birky *et al*, 2005). These factors that could have assisted genotype D *A. astaci* to differentiate significantly from the other genotypes before its introduction in Europe. With the knowledge accumulated so far on *A. astaci* and genotypes variability, it is possible that in the future genotypes could accumulate enough changes to be considered distinct species or subspecies.

The restriction digestion, the semi-nested and the amplification of genotype-specific regions assays developed in this study for the detection and discrimination of *A. astaci* known genotypes, were designed to target two different regions of *A. astaci* DNA: the nuclear DNA and the mtDNA. The three assays worked on pure cultures but presented different sensitivity when used on DNA extracted from infected crayfish from outbreaks and carriers' samples. As eukaryotic cells generally contain a high number of mitochondria (and therefore copies of mitochondrial genomic DNA) per cell (Luo *et al.*, 2011; Waugh, 2007), by targeting mtDNA the sensitivity of one of the assays improved allowing the detection of *A. astaci* genotypes from outbreaks and carriers samples. The difference in the representation of copy numbers between the ITS regions and the mtDNA in a DNA extraction has been investigated in this session to estimate the potential or amplification of a mtDNA target sequence versus a single copy nuclear sequence or the ITS regions.

ITS regions of the ribosomal DNA are the most employed DNA barcode used by the mycology and oomycete communities to identify and discriminate between species, including *Aphanomyces* (Bruns *et al.*, 1991; Cooke *et al.*, 2000; Diéguez-Uribeondo *et al.*, 2009; Lee and Taylor, 1992; Lévesque and De Cock, 2004; Robideau *et al.*, 2011; Schoch *et al.*, 2012). The number of copies of ITS in an individual organism has been estimated to be in the order of dozens to thousands, making this region easily targeted by conventional PCR methods (Álvarez and Wendel, 2003). The number of ITS regions in the nuclear genome of *A. astaci* calculated in this study correspond to 65, making this region a suitable target for conventional and high sensitive PCR methods for the detection of this pathogen in samples with low amount of DNA (Vrålstad *et al.*, 2009). Unfortunately, the intraspecific nucleotide variants noted between *A. astaci* isolates in the ITS regions were not consistent and could not discriminate one genotype from the other (section 3.2). Mitochondria are cellular organelles

used by eukaryotic cells to produce ATP and which possess their own genome (*i.e.* mtDNA) (Andersson *et al.*, 2003; Ernster and Schatz, 1981). Eukaryotic cells generally contain a high number of mitochondria and consequently a high number of copies of mtDNA in comparison to the nuclear DNA (Luo *et al.*, 2011; Waugh, 2007). In DNA extractions (*e.g.* DNA extractions used to obtain high molecular weight gDNA from pure *A. astaci* cultures subsequently utilised for WGS; DNA extractions from preserved crayfish tissues) from eukaryotic organisms (*e.g*. oomycetes), the total amount of DNA extracted will be a combination of nuclear DNA and a high proportion of mtDNA. Therefore, by targeting mtDNA instead of nuclear DNA in detection methods (*e.g.* PCR), the sensitivity of the test can be improved. The mean depth of coverage of the mtDNA calculated in this study for *A. astaci* D2 assembly resulted in 3105 reads for each nucleobase, against the mean depth of coverage of the whole genome which was 49 reads for nucleobase. This demonstrates the valuable potential of targeting this region versus targeting a single copy nuclear region.

4. Synthesis and recommendations

Aphanomyces astaci is a parasite of freshwater crayfish and involved in the decline of European indigenous crayfish species (Alderman *et al.*, 1984; Unestam, 1972) (section 1.2). Five genotypes (A, B, C, D, and E) have been detected and genetically characterised from pure cultures by random amplification of polymorphic DNA (RAPD-PCR) (Diéguez-Uribeondo *et al.*, 1995; Huang *et al.*, 1994; Kozubíková *et al.*, 2011) (section 1.2.5). Other fingerprinting tools that have been used to study crayfish plague include amplified fragment length polymorphism (AFLP) and microsatellites typing. These have been based on the same RAPD-PCR genotype classification (Grandjean *et al.*, 2014; Rezinciuc *et al.*, 2014) (section 1.3). Epidemiological studies of crayfish plague dependent on the distinction of *A. astaci* genotypes proved to be a useful tool to better understand the history and spread of this disease in Europe; at the same time differences in virulence between genotypes and between isolates belonging to the same genotype have been noted (Becking *et al.*, 2015; Lilley *et al.*, 1997a; Makkonen *et al.*, 2014; Makkonen *et al.*, 2012; Rezinciuc *et al.*, 2014; Viljamaa-Dirks *et al.*, 2013, 2016; Vrålstad *et al.*, 2014) (section 1.2.5) and being able to quickly and reliably identify *A. astaci* genotypes is an important step to study this disease and follow outbreaks in the field. The main objective of this study was to investigate the genetic diversity between *A. astaci* genotypes and to understand if genetic markers could be developed from these data. Ideally, these markers should bypass the culture methods (improvement from the RAPD-PCR and AFLP tools), differentiate between species of *Aphanomyces,* detect different genotypes and strains that can be present at the same time in the same sample, and to be used on samples with low concentration of the pathogen's targeted DNA (improvement from the microsatellite typing).

While confirming the identities of Cefas Oomycetes Culture Collection (OCC) isolates to be used in this study, phylogenetic analysis based on the internal transcribed spacer (ITS), the large subunit ribosomal DNA (LSU), and the cytochrome oxidase subunit I (COI) sequences were performed to detect nucleotide variants to be used as genotyping tools. The ITS and LSU analysis endorsed the efficacy of these markers to discriminate between species, but the resolution below species level for *A. astaci* isolates was not possible. The

phylogenetic analysis based on the COI sequences instead revealed that this gene presents nucleotide variants able to differentiate *A. astaci* genotype D isolates from the other genotypes. However, the other genotypes cannot be distinguished by these methods (section 3.2).

The advantages brought by next generation sequencing (NGS) and whole-genome sequencing (WGS) to population genetic studies resulted in cheap ways to discover genome-wide variations subsequently used to detect and develop new genetic markers (Schadt *et al.*, 2010; Studholme *et al.*, 2011). To develop new genotyping tools for the crayfish pathogen *A. astaci*, WGS was therefore employed to catalogue DNA single nucleotide variants and genotype-unique genomic regions that could be exploited as new phylogenetic markers. The knowledge gained from the WGS produced in this study of the *A. astaci* isolates held in the Cefas OCC and the *A. astaci* reference genome made publicly available by the Broad Institute, represented an important prerequisite to develop new reliable genotyping tools and to have a better understanding of the relationship between *A. astaci* genotypes (section 3.1). On one hand, the catalogued DNA SNVs and genotype-unique genomic regions were used to developed new genome-informed molecular markers to genotype *A. astaci*. On the other hand, WGS data highlighted similarities/differences between *A. astaci* genotypes and between the other *Aphanomyces* species sequenced in this study (section 3.6) while identifying loss of heterozygosity in one of *A. astaci* isolate's genome (section 3.1).

4.1 The new genotyping tools

The genotyping markers developed in this study were validated firstly on pure cultures and afterwards on historical samples (sections 3.3 and 3.4) and on a more recent well-characterised crayfish plague outbreak (section 3.5). These new genotyping methods outdo some of the shortcomings of the already available genotyping tools (AFLP, RAPD-PCR and microsatellites), implementing the detection of *A. astaci* genotypes from both pure cultures and infected material (sections 3.3, 3.4 and 3.5). They can be used to detect *A. astaci* from samples with a lower agent level (*i.e.* North American crayfish carriers) than the one necessary for the microsatellites technique (section

3.4.7). Moreover, the markers have been developed directly from known regions of the genome, and therefore PCR product obtained from the amplification of the targeted sequence can be directly sequenced, adding qualitative information to the test (sections 3.3, 3.4, and 3.5). The application of these markers in the laboratory, and the identification of *A. astaci* genotype involved in an outbreak, will add information to crayfish plague diagnosis, could help the competent authorities to track the source of the disease, and to identify the potential infected North American crayfish species introduced in the area: as specific North American crayfish generally carry specific *A. astaci* genotypes, knowing the genotype involved in an outbreak could indicate the presence/introduction of one of these carrier species (Diéguez-Uribeondo *et al.*, 1995; Huang *et al.*, 1994; Kozubíková, *et al.*, 2011; Lilley *et al.*, 1997a; Rezinciuc *et al.*, 2014). This information can then be used to search for the non-native species associated to the genotype, plan its eradication, and possibly trace the source of the infection (*e.g.* humans releasing ornamental trade animals in the wild, moving crayfish between basins). The genotyping tools developed in this study have been successfully applied to historical outbreaks and carriers' samples (section 3.4). Associating the occurrence of an outbreak in a geographical area and the *A. astaci* genotype related to it, could help to unveil the pattern of the disease in that particular area, adding valuable information to the study of genotype-host interactions and co-existence of the pathogen and host. Until now, this information has been collected indirectly by observing survival rates of European crayfish to crayfish plague epidemics, like in Turkey (Harlioğlu, 2004), or directly, though only rarely, by genotyping pure *A. astaci* cultures and conducting in-depth epidemiological surveys, like in Finland by Viljamaa-Dirks *et al*. (2013). Moreover, the application of the genotyping markers to screen infected crayfish stocks kept in captivity for repopulation purpose, could have a relevance in the prevention of new outbreaks. In particular, thanks to the extensive records and studies on crayfish plague in Europe, it is possible to distinguish two epidemic waves which devastated the European crayfish. The first plague epidemic ranges from the first introduction of the pathogen to the 1990's (Alderman, 1996), and the second one ranges from the 1980's onwards (Jussila *et al.*, 2016), including an overlapping period between the 1980's to the 1990's (Kozubíková *et al.*, 2008). This partition in waves is supported by *A. astaci* genotypes division, which

suggests that the first wave was caused by *A. astaci* genotype A, while the second wave was caused by *A. astaci* genotype B (Jussila *et al.*, 2016). Recent epidemiological surveys on crayfish plague in Finland demonstrate that European crayfish population weakened by crayfish plague epizootic events occurred during the first wave can carry genotype A as a form of persistent chronic infection. Some of these populations subsequently faded due to new arrival (second wave) of North American crayfish carrying *A. astaci* genotype B (Viljamaa-Dirks *et al.*, 2013). This hypothesis is supported by the studies of Makkonen *et al.* (2012b) indicating an increased *A. astacus* resistance against genotype A due to co-evolution. Therefore, if no *A. astaci*-free crayfish individuals are available for repopulation purpose of an area known to have been previously infected with a particular genotype, the reintroduction of crayfish in the area should be undertaken with animals infected with the same genotype. The old population that might still be living in the area might be already adapted to the genotype and the reintroduction of the same genotype might not result in serious outbreaks. However, diagnostic and genotyping tools for *A. astaci* have a limited use to prevent outbreaks in the wild as human activities (*e.g.* release of ornamental animals in the wilds, moving animals between areas, moving infected equipment between areas) are uncontrollable and represent one of the biggest threat to local crayfish populations. The only available tool to prevent new introductions and spreads is actively educating and sensitising the local community to the subject. The genotyping tools developed in this study are new tools and to fully understand their potential they should be tested alongside the already established genotyping techniques. However, their features clearly demonstrate the extreme value of these new genotyping markers, which can be used on pure cultures, on recent infected material or on historical samples and therefore are a useful tool in aid of the identification of *A. astaci* in epidemiological studies.

Besides the new genotyping techniques developed here, a note has to be made to the value of the isolation and upkeeping of pathogens in pure culture. Without pure cultures, this and other numerous studies on the characterisation of the properties of *A. astaci* isolates would have not been possible. Being able to identify *A. astaci* genotypes from fresh and preserved materials is an achievement to pursue, especially for a quick diagnosis during an outbreak or a

comprehensive epidemilogical study, but the importance of the data and the exploitation of the live culture (*i.e. in vivo/in vitro* infection studies) that can be achieved only by using a pure culture should not be underestimated. Therefore, maximum effort should be applied to the isolation of the pathogen.

4.2 *Aphanomyces* genomes general outlook

The data obtained by the application of WGS to all known *A. astaci* genotypes and to the other three *Aphanomyces* species highlighted similarities and differences between *A. astaci* genotypes and between the other *Aphanomyces* species sequenced in this study (sections 3.1 and 3.6). For example, while genotyping *A. astaci* pure cultures held in Cefas OCC with the RAPD-PCR approach, an atypical RAPD-PCR profile for *A. astaci* isolate 457 was observed (section 3.1). Previously, this isolate was ascribed to genotype B. A deep investigation into this anomaly using WGS data interestingly showed LOH in this isolate in the same genome region targeted by the RAPD-PCR primer, and therefore influencing the isolate's RAPD-PCR profile. LOH have been previously observed in oomycetes kept in laboratory conditions for a long period of time, similarly as *A. astaci* isolate 457 (Dobrowolski *et al.*, 2002; Hu *et al.*, 2013; Lamour *et al.*, 2012; Shrestha *et al.*, 2014). LOH has been associated with mating strategy changes during asexual growth, typical in laboratory condition (Shrestha *et al.*, 2014). The meaning of LOH is still not clear in oomycetes, but it could reduce and remove dominant alleles and traits to favour recessive alleles contributing to the adaptation to new environments and loss of pathogenicity (Jiang *et al.*, 2013; Kasuga *et al.*, 2016; Lamour *et al.*, 2012), characteristics that could be an advantage for *A. astaci* to successfully coexists with the European crayfish hosts (Jussila *et al.*, 2016).

Further studying the data obtained from the WGS considering the SNVs phylogenetic analysis, *A. astaci* isolate 457 was decisively included between the genotype B isolates'. WGS data analysed by dnadiff also highlighted similarities and differences between *A. astaci* genotypes and between the other *Aphanomyces* species sequenced in this study (section 3.6). By genome comparison, between 9 to 13 % of sequence differences were sufficient to discriminate between the aquatic *Aphanomyces* species used in this study,

while isolates belonging to the same species shared at least 95 % of sequences similarities. *A. astaci* isolates belonging to the same genotype, instead, shared at least 99 % of sequence similarity. Based on this result, while *A. invadans* isolates can be ascribed to a single clone/genotype world-wide distributed by commercial trades of aquatic animals confirming Lilley *et al*. (1997b) theory, *A. frigidophilus* isolates presented enough genomic variation to indicate the presence of different clones/genotypes, feature so far not investigated in this species. Moreover, *A. invadans*-like presented high (90 %) sequence similarity to *A. frigidophilus*, indicating that these opportunistic/saprotrophic species share more genome sequence similarity among them rather than with animal parasitic *Aphanomyces* species.

4.3 *Aphanomyces astaci* genome: beyond genotyping tools

The knowledge and the data obtained by the application of WGS to all known *A. astaci* genotypes goes beyond the development of genotyping tools. For instance, the dnadiff analysis of *A. astaci* genome assemblies generated in this study (section 3.6) pointed out sequence differences between isolates belonging to different genotypes and between isolates belonging to the same genotype. These genetic variabilities could harbour important features or information regarding the differences in virulence noted between *A. astaci* genotype A and B during infection trials with European and North American crayfish (Aydin *et al.*, 2014; Becking *et al.*, 2015; Jussila *et al.*, 2013; Makkonen *et al.*, 2012; Makkonen *et al.*, 2014; Viljamaa-Dirks *et al.*, 2016). This feature is an important aspect to investigate further and the new genome information could shed light on the genotype differences, setting up the future framework in this field. For example, carbohydrate-active enzymes (CAZys), involved in breaking down of chitin, are putative virulence factors of those animal pathogenic oomycetes which parasitise organisms with cuticles, like crayfish, targeted by *A. astaci* (Misner *et al*., 2016; Roer and Dillaman, 1984). This group of enzymes includes chitinase genes, *A. astaci* virulence factors, which have already been shown to possess sequence variability (Makkonen *et al*., 2012). The nucleotide variations of these genes were investigated by Makkonen *et al.* (2012) in an attempt to develop genotyping tools. By analysing four genotypes

(A, B, C, and D), the authors identified three chitinases groups (CHI1 to 3) and identified not only nucleotide variations between genes, but also discovered that some of the genes characteristic of one genotype were not shared with other genotypes, emphasising the evolutionary importance of these genes in *A. astaci* (Makkonen *et al*., 2012). As these important virulence factors, which might be expressed during infection (Hochwimmer *et al.*, 2009), show variability between *A. astaci* genotypes, genetic differences in other not yet analysed pathogenicity factors could indeed be present in the genotypes and could be the base of the already noted virulence differences (Aydin *et al.*, 2014; Becking *et al.*, 2015; Jussila *et al.*, 2013; Makkonen *et al.*, 2012; Makkonen *et al.*, 2014; Viljamaa-Dirks *et al.*, 2016). Looking at the genome of each genotype with more comprehensive tools could lead to the identification of these virulence factors and their differences. As an example, by predicting and annotating the genes from the genomes of each *A. astaci* known genotype, it would be possible to identify orthologs [genes present in different species, which have evolved from a common ancestral gene (Fitch, 1970)] of putative pathogenicity factors already characterised in plant pathogenic oomycetes like *Phytophthora infestans* (Haas *et al*., 2009), animal pathogenic oomycetes like *Saprolegnia parasitica* (Jiang *et al.,* 2013) and *Lagenidium giganteum* (Quiroz Velasquez *et al.,* 2014), the facultative decapod parasite *Achlya hypogyna* (Misner *et al*., 2015), and the non-pathogenic saprobe *Thraustotheca clavata* (Misner *et al*., 2015). Some of these putative pathogenicity factors, for which *A. astaci* could have orthologs, include for example elicitin and elicitin-like effector proteins. This group of proteins were firstly detected in plant pathogenic oomycetes like *Phytophthora* and *Pythium* (Jiang *et al*., 2006; Panabieres *et al*., 1997) and recently detected also in animal pathogenic oomycetes like *Saprolegnia* (Jiang *et al*., 2013). They are involved in extracellular lipid binding and transfer, and have been shown to trigger the defence response in plants (Jiang *et al*., 2006; Panabieres *et al*., 1997). Other putative pathogenicity factors identified only in the saprolegniales like *A. hypogyna*, *S. parasitica,* and *T. clavata,* and the peronosporales *L. giganteum*, but absent in plant pathogenic oomycetes, are the glycoside hydrolases belonging to the family 20 (GH20), which bind carbohydrates during hosts cuticles degradation (Misner *et al*., 2015; Olivera *et al.,* 2016; Quiroz Velasquez *et al.,* 2014). Virulence factors abundantly present in plant pathogenic oomycetes are the “crinkler” (or CRN, from crinkling and

necrosis) and the RxLR (arginine-x-leucine-arginine, with x as any amino acid) protein families, which are host-translocated effectors responsible for the suppression of the plant defence system and which cause plant cells death (Birch *et al.,* 2009; Bos, 2007; Lévesque *et al.*, 2010). These effectors translocate from the oomycete to the targeted host cells only if the protein present particular motifs: the LxLFLAK (leucine-x-leucine-phenylalanine-leucine-alanine-lysine, with x as any amino acid) motif for the CRN effectors; the RxLR motif for the homonym effectors (Dou *et al*., 2008; Schornack *et al*., 2010; Stam *et al*., 2013; Whisson *et al.*, 2007). These pathogenicity factors seem to lack from the genome of animal pathogenetic oomycetes, except for *L. giganteum* which, however, does not express these pathogenicity factors during infection (Olivera *et al.,* 2016). *S. parasitica* presents instead a similar family of proteins containing a terminal RxLR, called SpHtp1 (or *S. parasitica* host-targeting protein 1) (Jiang *et al*., 2013; van West *et al.,* 2010; Wawra *et al.*, 2012). *In vitro* studies involving the infection of fish cells lines (rainbow trout cells) with one of the proteins belonging to SpHtp1 family, have shown that the protein could translocate into the cells and therefore can be considered a pathogenicity factor (van West *et al.,* 2010; Wawra *et al*, 2012). Once the genome of each *A. astaci* genotype has been annotated, it would be possible to locate the presence (or absence) of pathogenicity factors in each genotype and to directly compare the annotations to identify sequence differences and unique features of each genotype. Moreover, the difference in virulence and pathogenicity between *A. astaci* genotypes and isolates against European and North American crayfish could be further investigated in infections trials by analysis of either the whole transcriptome or by targeting important features (*e.g.* genes and proteins) detected by the genome prediction and annotation analysis. Examining the transcriptome of single isolates during different stages of their life cycle, (*e.g.* vegetative vs infection stages), would identify genes that are being actively expressed at the time and give a comprehensive view of which genes are used by *A. astaci* during the infection process. By comparing different isolates and genotypes, it would be possible to look deeply into the virulence differences between genotypes not only by a genetic point of view (*i.e.* gene prediction and annotation) but also by a transcriptomic point of view, and it would be possible to identify which genes are up or down regulated.

Analysing the genome of *A. astaci* could also point out paticular sectered proteins (secretome) that could be investigated as potential candidate for vaccine development. In contrast to vertebrates, invertrebrates as crayfish lack the specific immune system mediated by lymphocytes and antibodies and thier immune system is usually described as non-specific, or innate, without immune memory (Rowley and Pope, 2012). Recent studies on insect, however, shown that invertebrates may possess a “specific immune priming” similar to the one of vertrebrates, which can lead to increased survival rates and trans-generational (transmitted to offspring) enahnced immune response (Little *et al*., 2003; Pham *et al*., 2007; Roth *et al*., 2010). However, the underlying mechanisms involved in the immune priming have not been fully investigated and seems to be limited to arthropods and not crustaceans (Rowley and Pope, 2012). Vaccination of crustacean (shrimps) against fatal shrimp pathogens, like the white-spot syndrome virus (WSSV), have been investigated in several studies, which highlighted increased survival rates following vaccination with inactivated virus, viral envelope proteins, DNA, and RNA vaccines (Kim *et al.,* 2007; Li *et al.,* 2010; Singh *et al*., 2005; Wei and Xu, 2009). However, the protection gained by the crustacean against the disease was temporarly limited. For example, following the oral delivery of dsRNA based vaccines, Sarathi *et al*. (2008) noted 68 % of survival after 30 days after infection with WSSV, while Li *et al.* (2010) found that the survival rate of shrimps vaccinated at the same time with recombinant viral proteins decreased from 50 % to 20 % when infected a week apart, implying loss of protection. Even if the vaccination of shrimps against WSSV only reduces the mortalities against this virus, it is an advantage in shrimp aquaculture to prevent income loss due to WSSV outbreaks. Overall, the development of vaccines agaist *A. astaci* to immunise European crayfish is possible. However, the 85 % of world crayfish production for human consumption is composed by red claw crayfish, (*Procambarus clarkii*), which is an asyntomatic carrier of *A. astaci*, while European crayfish are not farmed for human consumption but only for repopulation purpose (Holdich, 1993). Even if it is still unclear how the mechanisms of immunisation work in crustacean, how long this immunisation lasts for, and if it is actively passed onto offspring, vaccinating European crayfish against the crayfish plague pathogen is a tempting possibility, and more effort should be put in place to evaluate this tool. For esample, a vaccination program to boost crayfish immune response before

the repopulation of an infected area could be a useful tool to prevent new deadly outbreaks upon the release of the animals in the wild.

4.4 *Aphanomyces* species genomes analysis: future work

The genome, secretome, and pathogenicity of plant pathogenic oomycetes has been deeply and profusely studied from a genomic and transcriptomic point of view (Kamoun *et al*., 2015; Thines and Kamoun, 2010). Recent analysis of the genome and transcriptome of animal pathogenic oomycetes like *Saprolegnia parasitica* (Jiang *et al.,* 2013), *Lagenidium giganteum* (Quiroz Velasquez *et al.,* 2014), and *Achlya hypogyna* (Misner *et al*., 2015), and the non-pathogenic saprobe *Thraustotheca clavata* (Misner *et al*., 2015) highlighted differences and similarities between plant and animals pathogenetic oomycetes and saprotrophic opportunistic oomycetes, while revealing the potential of comparative genomic analysis between vast spectrum of species, and closing the gap of information between animal and plant pathogenic oomycetes (Jiang *et al.,* 2013; Misner *et al*., 2015; Olivera *et al*., 2016; Quiroz Velasquez *et al.,* 2014; van West *et al.,* 2010; Wawra *et al*, 2012).

The *Aphanomyces* genus comprises a broad range of oomycetes species adapted to parasitise plants (*e.g. A. euteiches*), animals (*e.g. A. invadans* and *A. astaci*), and saprotrophic/opportunistic species (*e.g. A. stellatus*) (Diéguez-Uribeondo *et al.*, 2006). By applying comparative genomics on the genome of these different broad range of species and lifestyles, it could be possible to understand the evolution of parasitism and adaptation to different environments. Until recently, only *A. euteiches* (plant pathogen) transcriptome, *A. invadans* (fish pathogen) genome, and *A. astaci* (crayfish pathogen) genotype D genome, were available to be compared (Gaulin *et al*., 2008; *A. astaci* GenBank assembly accession: GCA_000520075.1; *A. invadans* GenBank assembly accession number: GCA_000520115.1). In the present study, two other *Aphanomyces* species with different lifestyles have been sequenced: *A. invadans*-like, a saprotrophic/opportunistic species, and *A. frigidophilus*, a species isolated from animals, grouping with animal pathogenic *Aphanomyces* in phylogenetic studies (sections 3.2.3, 3.2.4, and 3.2.5; Diéguez-Uribeondo *et al.*, 2009), but still unclear if saprotrophic/opportunistic as infection studies failed to determine its pathogenicity. Comparing the predicted genes between the

species mentioned above, with reference to pathogenicity factors and metabolic pathways, could highlight the evolution of parasitism and host selection in the *Aphanomyces* genus. The dnadiff analysis conducted in the present study (3.6) already gives clues on the similarity/differences between species and lifestyles. For examples, *A. frigidophilus* and *A. invadans*-like share more sequence similarities than with the other animal pathogenic oomycetes. Therefore, by doing gene prediction, annotation, and genome comparative analysis of the genomes available, it could be possible to better understand the evolution and speciation of this genus.

While plant pathogenic oomycetes shift (from a saprotrophic to a pathogenic lifestyle) has been related to the expansion of effectors and pathogenicity factors (Olivera *et al.,* 2016; Schornack *et al*., 2010), for animal pathogenic oomycetes this shift has been characterised by the expansion of potential virulence factors in the form of secreted proteases, lectins and toxins acquired from animals or animal pathogens by horizontal gene transfer (Jiang *et al*., 2013). As for *S. parasitica*, the expansion of protease genes has also been recorded in *A. euteiches*, which play a major role in *A. euteiches* pathogenicity (Gaulin *et al*., 2008). The broad host range, the specialisation, and the successful pathogenicity strategies of oomycetes as a class has been associated to the expansion of their pathogenicity factors, which genes are located in genome regions characterised by low level of genome conservations and therefore subjected to a high evolutionary rate (Jiang *et al.,* 2008; Soanes and Talbot, 2008; Raffaele *et al*, 2010).

The animal pathogenic oomycetes studied so far, present some unique pathogenicity factors, not shared with plant pathogenic oomycetes. One example are the glycoside hydrolases belonging to the family 20 (GH20) involved in during hosts cuticles degradation, found in *S. parasitica*, *L. giganteum, A. hypogyna*, *A. invadans,* and *A. astaci* (Misner *et al*., 2015; Olivera *et al.,* 2016; Quiroz Velasquez *et al.,* 2014), while *A. euteiches* present GH5 and GH81 families involved in degradation of plant cellulose and callose (Gaulin *et al*., 2008). Moreover, on the top of the GH20 proteins involved during hosts cuticles degradation, *S. parasitica* oddly presents a small repertoire of enzymes to degrade pectin (constituent of plant cell walls) indicating that *S. parasitica* could live as a saprotrophic species when not infecting a fish host

(Jiang *et al*., 2013). In a saprotrophic/opportunistic species like *A. frigidophilus* and *A. invadans*-like, both GH characteristics of animal and plant pathogenic families could be present to sustain life during the saprotrophic and parasitic life stages. Pathogenicity factors shared to some extent with plant pathogenic oomycetes are the RxLR (arginine-x-leucine-arginine, with x as any amino acid) protein families, which are host-translocated effectors responsible for the suppression of the plant defence system and which cause plant cells death (Birch *et al.,* 2009; Bos, 2007; Lévesque *et al.*, 2010). *S. parasitica* presents a similar small family of proteins containing a terminal RxLR, the SpHtp1 (or *S. parasitica* host-targeting protein 1), which can translocate into fish cells in *in vitro* studies and are considered pathogenicity factors (Jiang *et al*., 2013; van West *et al*., 2010; Wawra *et al*., 2012). While *S. parasitica* genome mining did not identify any CRN pathogenicity factors, *A. euteiches* (plant pathogen) transcriptome revealed the presence of several CRN sequences, but no RxLR (Gaulin *et al.,* 2008). An in-depth analysis of the *Aphanomyces* species genomes sequenced in the present study could reveal the presence (or absence) of similar effectors. At present, bioinformatical searches of these pathogenicity factors in the genome and transcriptome are based on sequence homology. However, the genes encoding for these effectors are located in variable regions and are often subjected to strong positive selection, therefore eluding *in silico* analyses (Gaulin *et al*., 2008; Raffaele *et al*., 2010; Win and Kamoun, 2008). The absence of effectors in bioinformatics analyses could be a bias of the technique used and cannot completely indicate that they are missing from the genome. Beyond pathogenicity factors, an insight into the metabolic pathways could indicated the differences between pathogenic and non-pathogenic *Aphanomyces.* For example, comparing the genome of *S. parasitica* to plant pathogenic oomycetes, revealed a reduction of metabolic pathways involved in assimilation of inorganic nitrogen and sulfur, possibly derived from the adaptation of *S. parasitica* to a parasitic lifestyle on animal tissues which are rich in proteins and ammonium (Jiang *et al*., 2013). Similar events have also been recorded in obligate parasitic oomycetes and fungal plant pathogens (McDowell*,* 2011). Therefore, by comparing *Aphanomyces* species with different lifestyles, it could be possible to identify differences in metabolisms and comprehend the evolution of parasitism and the adaptation to different hosts for this genus.

5. Appendix

5.1 Section 2.1.8

Table 5.1 Cefas historical carrier samples from infected North American crayfish (signal crayfish - *P. leniusculus*). North American crayfish samples had been from various locations in water courses in England and Wales. The samples tested positive for *A. astaci* at Cefas by qPCR (Tuffs and Oidtmann, unpublished data), following Vrålstad *et al.* (2009). Agent level classification follows Vrålstad *et al.* (2009).

Population	Location	Animal number	Ct mean	DNA quantity (ng)	Agent level
III	River Lee	13-telson	31.93804	0.000395	A2
III	River Lee	28-cuticle	35.0575	0.000141	A2
IV	River Ash	30-telson	36.2372	0.000334	A2
VII	Tetbury Avon	26-telson	37.4973	0.000365	A2
VIII	Hamps	6-telson	37.0562	0.000082	A1
IX	Bently Brook	6-cuticle	28.3677	0.005057	A3
XIV	Bachaway	16-cuticle	36.0431	0.000023	A1
XIV	Bachaway	19-cuticle	34.9827	0.000022	A1
XV	River Wid	2-cuticle	31.2817	0.000173	A2
XV	River Wid	8-cuticle	31.2096	0.000252	A2
XVI	River Wharfe	2-cuticle	34.1255	0.000074	A1
XVI	River Wharfe	8-cuticle	33.0347	0.00009	A1
XVII	River Evenlode	6a1-cuticle	36.6282	0.000044	A1
XVII	River Evenlode	9a2-cuticle	34.6151	0.000094	A1
XVIII	River Thame	14a2-cuticle	32.6632	0.000245	A2
XVIII	River Thame	16a1-cuticle	32.9202	0.000248	A2

Table 5.2 Cefas historical outbreaks samples from European crayfish (white clawed crayfish - *A. pallipes*). The samples had tested positive for *A. astaci* at Cefas by PCR using primers 42 and 640 and sequencing of PCR product (OIE, 2016a).

Outbreak number	Abbreviation	Sample number	Outbreak year
PM28325	O1	1.2	2014
PM28325	O1	1.4	2014
PM28465	O2	1.2	2014
PM21018	O3	1	2010
PM21018	O3	2	2010
PM21018	O3	3	2010
PM19790	O4	1	2009
PM19790	O4	2	2009
PM19790	O4	6	2009
PM19790	O4	7	2009
PM19955	O5	1.1	2009
PM19955	O5	2.1	2009
PM19955	O5	3.1	2009
PM-M17120	O6	1	2007
PM-M17120	O6	2	2007

Table 5.3 Italian outbreak samples. Genomic DNA extracted from the cuticle of white clawed crayfish (*Austropotamobius pallipes*) or from pure of *A. astaci* isolated during the outbreak.

Type of sample	Sample number	Ct mean
Cuticle	17/ITT/14.11	22.6
	17/ITT/14.16	21.94
	69 /ITT/14.1	15.28
	69 /ITT/14.7	19.61
Pure culture	117/ITT/14.α23	-
	17/ITT/14.25βbis	-
	17/ITT/14.2bis	-

5.2 Section 2.4.2

Table 5.4 *Aphanomyces* species and *Saprolegnia parasitica* sequences retrieved from NCBI Nucleotide database and used for COI phylogenetic analysis. *S. parasitica* was used as outgroup. **A. astaci* isolate APO3 belongs to genotype D (J. Diéguez-Uribeondo, personal communication, May 2014).

Oomycetes species	Isolate ID	GenBank accession number
*A. astaci**	AP03	SRR957817
A. cladogamus	CBS10829	HQ708186
A. cochlioides	CBS47771	HQ708188
A. euteiches	CBS15473	HQ708192
A. iridis	CBS52487	HQ708194
A. laevis	CBS47871	HQ708195
S. parasitica	CBS113187	HQ709046

Table 5.5 *Aphanomyces* species and *Saprolegnia parasitica* sequences retrieved from NCBI Nucleotide database and used for ITS phylogenetic analysis. *S. parasitica* was used as outgroup. Bold: reference sequences as described in Diéguez-Uribeondo *et al.* (2009).

Oomycetes species	Isolate ID	GenBank accession number
A. astaci	**L1**	**AY310501**
A. cladogamus	**SAP355**	**FM999228**
A. cochlioides	**94**	**AY353911**
A. euteiches	**SAP368**	**FM999226**
A. frigidophilus	**SAP263**	**FM999233**
A. helicoides	**CBS210.82**	**AY310496**
A. invadans	**SAP307**	**FM999229**
A. iridis	**SAP356**	**FM999227**
A. laevis	**SAP366**	**FM999237**
A. repetans	**Se**	**AY683897**
A. salsuginosus	NJM0802	AB510349
A. stellatus	**CBS568.67**	**AM947029**
Aphanomyces sp.	84-1240	AF396683
S. parasitica	NJM9880	AY455776

Table 5.6 *Aphanomyces astaci* and *A. invadans* sequences retrieved from NCBI Nucleotide database and used for *A. astaci* ITS phylogenetic analysis. *A. invadans* was used as outgroup. Bold: *A. astaci* isolates used as genotype references.

Oomycetes species and genotype (for *A. astaci*)	**Isolate ID**	**GenBank accession number**	**Genotype reference**
***A. astaci* genotype A**	**L1**	**AY310501**	**Oidtmann *et al.* (2004)**
A. astaci genotype A	L1	AY683895	Royo *et al.* (2004)
A. astaci genotype A	EviraK071/99	GU320232	Makkonen *et al.* (2011)
A. astaci genotype A	L1	GU320248	
***A. astaci* genotype B**	**P1**	**AM947023**	**Vrålstad *et al.* (2009)**
A. astaci genotype B	Ho	AY683893	Royo *et al.* (2004)
A. astaci genotype B	FDL457	AY310500	Oidtmann *et al.* (2004)
A. astaci genotype B	M96/1	AY310499	
A. astaci genotype B	EviraK104/98	GU320233	Makkonen *et al.* (2011)
A. astaci genotype B	Evira4426/03	GU320247	
A. astaci genotype B	UEF7208	GU320214	
A. astaci genotype B	Evira3697/03	GU320236	
A. astaci genotype B	Evira7862/03	GU320237	
A. astaci genotype B	Evira0905	GU320231	
A. astaci genotype B	UEF8866-2	GU320225	
A. astaci genotype B	UEF7203	GU320238	
A. astaci genotype B	UEF7204	GU320239	
A. astaci genotype B	UEF8140-2	GU320221	
A. astaci genotype B	Evira3697/03	GU320229	
***A. astaci* genotype C**	**Kv1**	**AY683894**	**Royo *et al.* (2004)**
***A. astaci* genotype D**	**Pc**	**AY683896**	**Royo *et al.* (2004)**
A. astaci genotype D	APO3	XR717100	J. Diéguez-Uribeondo, Personal communication May 2014
***A. astaci* genotype E**	**Evira4805b/10**	**JF827153**	**Kozubíková *et al.* (2011)**
A. invadans	SAP307	FM999229	-

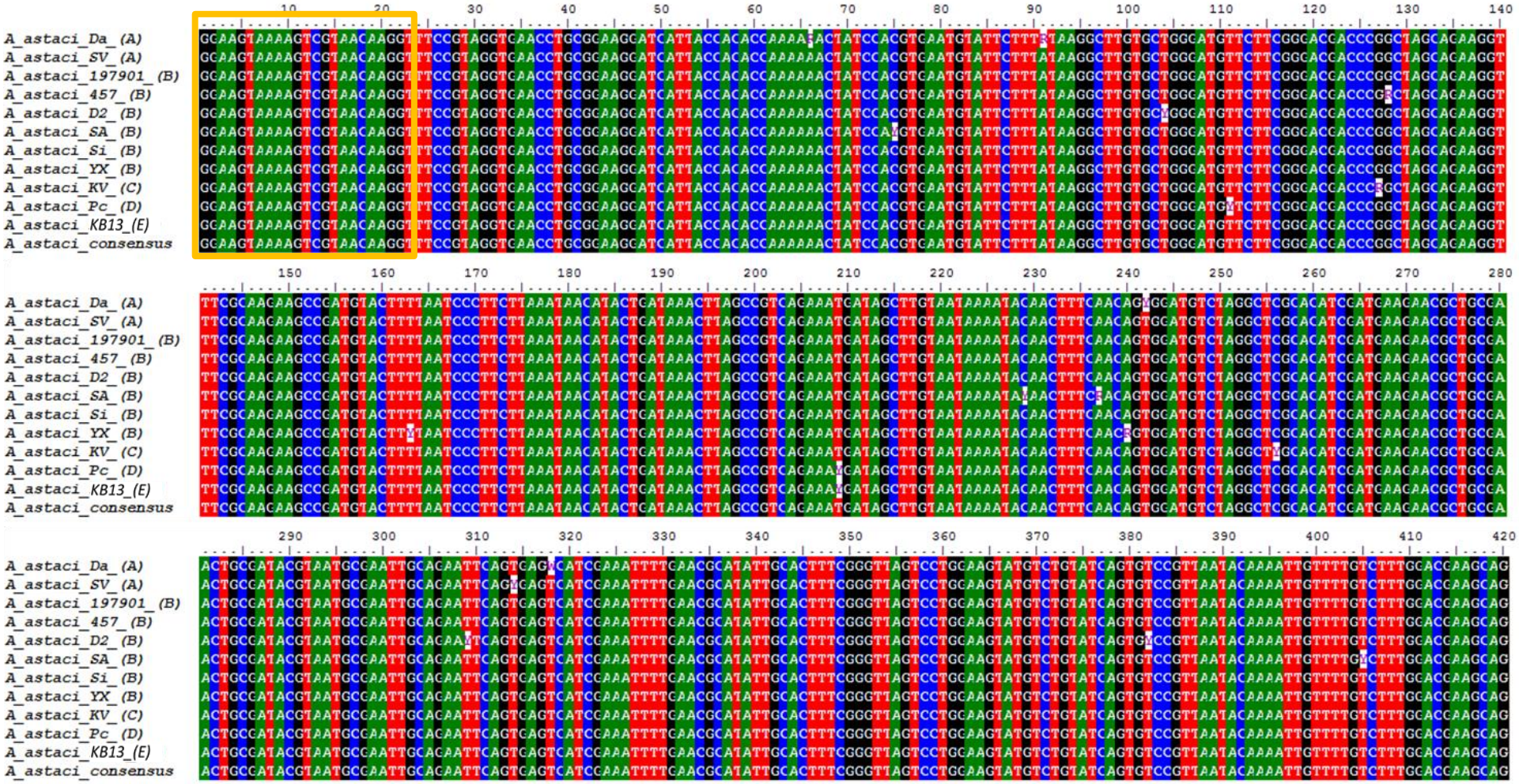
Table 5.7 *Aphanomyces* species and *Saprolegnia parasitica* sequences retrieved from NCBI Nucleotide database and used for LSU phylogenetic analysis. *S. parasitica* was used as outgroup. - Unknown information.

Oomycetes species and genotype (for *A. astaci*)	**Isolate ID**	**GenBank accession number**	**Genotype reference**
A. astaci genotype A	uv	AF235940	Petersen and Rosendahl (2000)
A. astaci genotype B	Ho	AF218197	Leclerc *et al.* (2000)
A. astaci genotype D	AP03	XR717097	J. Diéguez-Uribeondo, personal communication, May 2014
A. astaci genotype D	AP03	XR717100	
A. cladogamus	CBS108.29	HQ665056	-
A. cochlioides	CBS477.71	HQ665241	-
A. euteiches	CBS156.73	HQ665132	-
A. invadans	NJM9701	XM008883388	-
A. iridis	CBS524.87	HQ665248	-
A. laevis	CBS478.71	HQ665242	-
A. repetans	CBS126887	HQ395671	-
A. salsuginosus	ATTC MYA-4775	JQ070132	-
A. sinensis	ATTC MYA-4825	JQ070142	-
A. stellatus	GF	JN662487	-
Aphanomyces sp.	CBS583.85	HQ665276	-
S. parasitica	CBS302.56	HQ665197	-

5.3 Section 3.2.3

Table 5.8 List of differences in ITS cloned sequences and mutation sites. For *A. astaci* Si and *A. invadans* NJM0002 no base changes were detected in the cloned sequences.

Species-isolate ID-clone number	Differences between clones (nt)	Variable site consensus>mutation
A. astaci 197901-12	1	774A>G
A. astaci 457-4	2	128G>A, 625A>G
A. astaci 457-5	2	422A>G, 590G>A
A. astaci 457-6	1	518A>G
A. astaci D2-2	2	382T>C, 487T>A
A. astaci D2-4	2	114T>C, 676T>C
A. astaci D2-5	3	554T>C, 616T>A, 679T>C
A. astaci D2-6	2	309T>C, 487T>A
A. astaci Da-6	3	545T>C, 649T>C, 709A>G
A. astaci Da-7	2	91A>G, 318T>A
A. astaci Da-8	2	66A>G, 242T>C
A. astaci KV-4	1	256C>T
A. astaci KV-35	1	127G>A
A. astaci KB13-6	2	776G>A, 777A>T
A. astaci KB13-13	1	209T>C
A. astaci Pc-3	1	209T>C
A. astaci Pc-5	1	111T>C
A. astaci SA-2	3	237A>G, 405T>C, 503T>C
A. astaci SA-3	1	229C>T
A. astaci SA-6	1	75C>T
A. astaci SV-3	1	314T>C
A. astaci SV-4	1	514T>gap
A. astaci YX-1	1	629G>A
A. astaci YX-5	3	240A>G, 479A>G, 584C>T
A. astaci YX-7	1	163T>C
A. astaci YX-10	1	598A>G
A. frigidophilus AP5-2	2	173T>A, 415G>A
A. frigidophilus RP1-1	1	551G>A
A. frigidophilus RP1-3	1	398T>C
A. frigidophilus RP1-6	1	189T>A
A. frigidophilus RP1-7	1	481A>G
A. frigidophilus RP2-1	3	139A>G, 398T>C, 500T>C
A. frigidophilus RP2-7	2	84A>T, 127A>G
A. invadans GWR-3	2	596T>C, 722A>T
A. invadans GWR-11	1	647A>G
A. invadans NJM8997-2	2	440A>G, 621T>C
A. invadans NJM8997-3	2	287A>T, 456A>G
A. invadans NJM9030-2	3	1G>A, 2G>T, 59A>G
A. invadans NJM9030-4	1	171A>G
A. invadans NJM9030-5	1	220T>C
A. invadans NJM9701-6	1	625A>G
A. invadans NJM9701-8	1	627A>G
A. invadans-like NJM9510-9	1	308T>C
A. invadans-like NJM9510-33	6	157A>C, 403T>C, 435A>C 580A>G, 584G>T, 701A>T


A_astaci_Da_(A)
A_astaci_SV_(A)
A_astaci_197901_(B)
A_astaci_457_(B)
A_astaci_D2_(B)
A_astaci_SA_(B)
A_astaci_Si_(B)
A_astaci_YX_(B)
A_astaci_KV_(C)
A_astaci_Pc_(D)
A_astaci_KB13_(E)
A_astaci_consensus

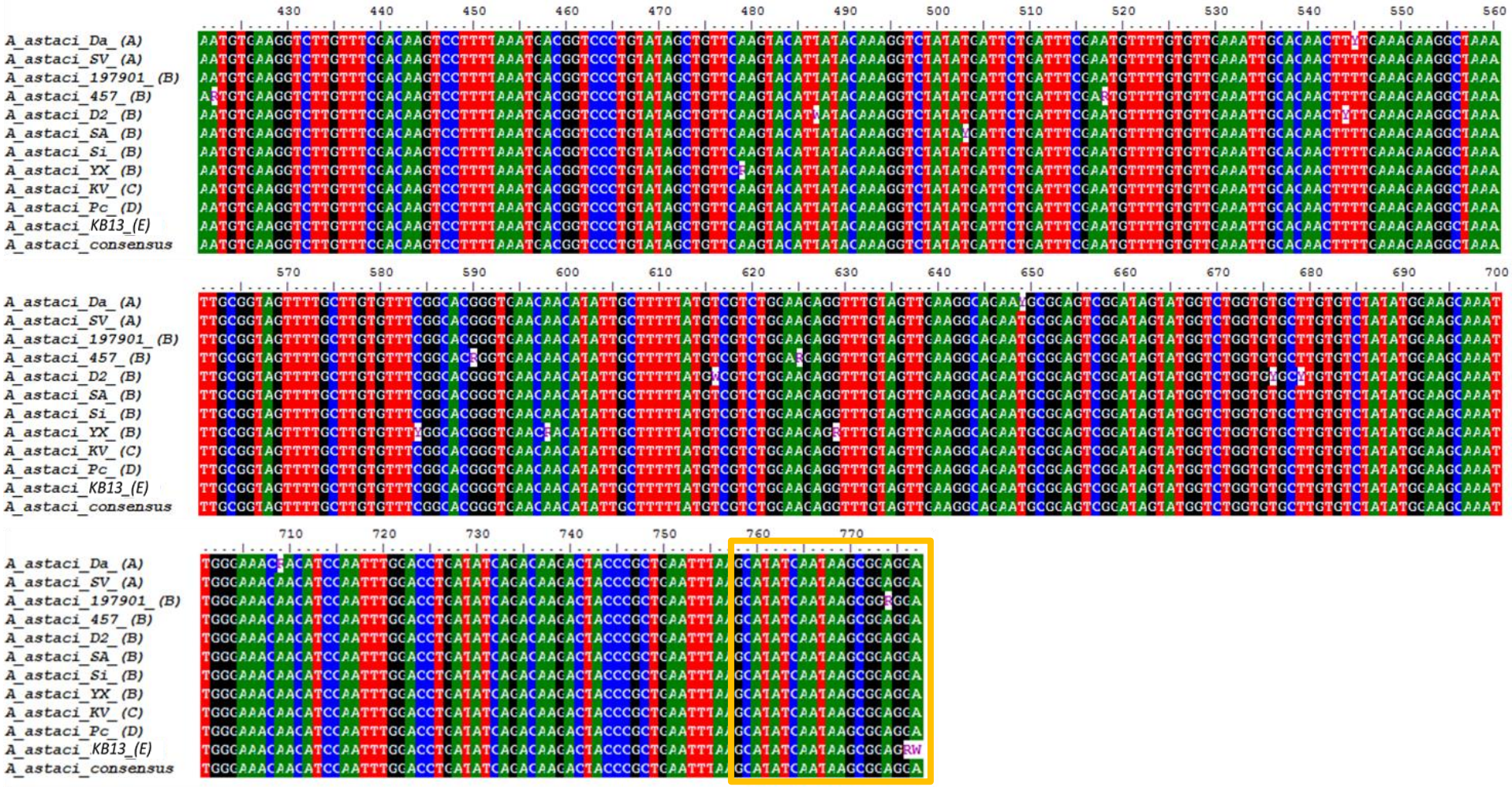


Figure 5.1 *A. astaci* ITS consensus sequences showing variable sites. Consensus sequences were generated with BioEdit: for each isolate, sequences were obtained combing the respective clones; "*A. astaci* Consensus" sequence was generated by combing all isolates. Yellow boxes: primers sites.

5.4 Section 3.2.4

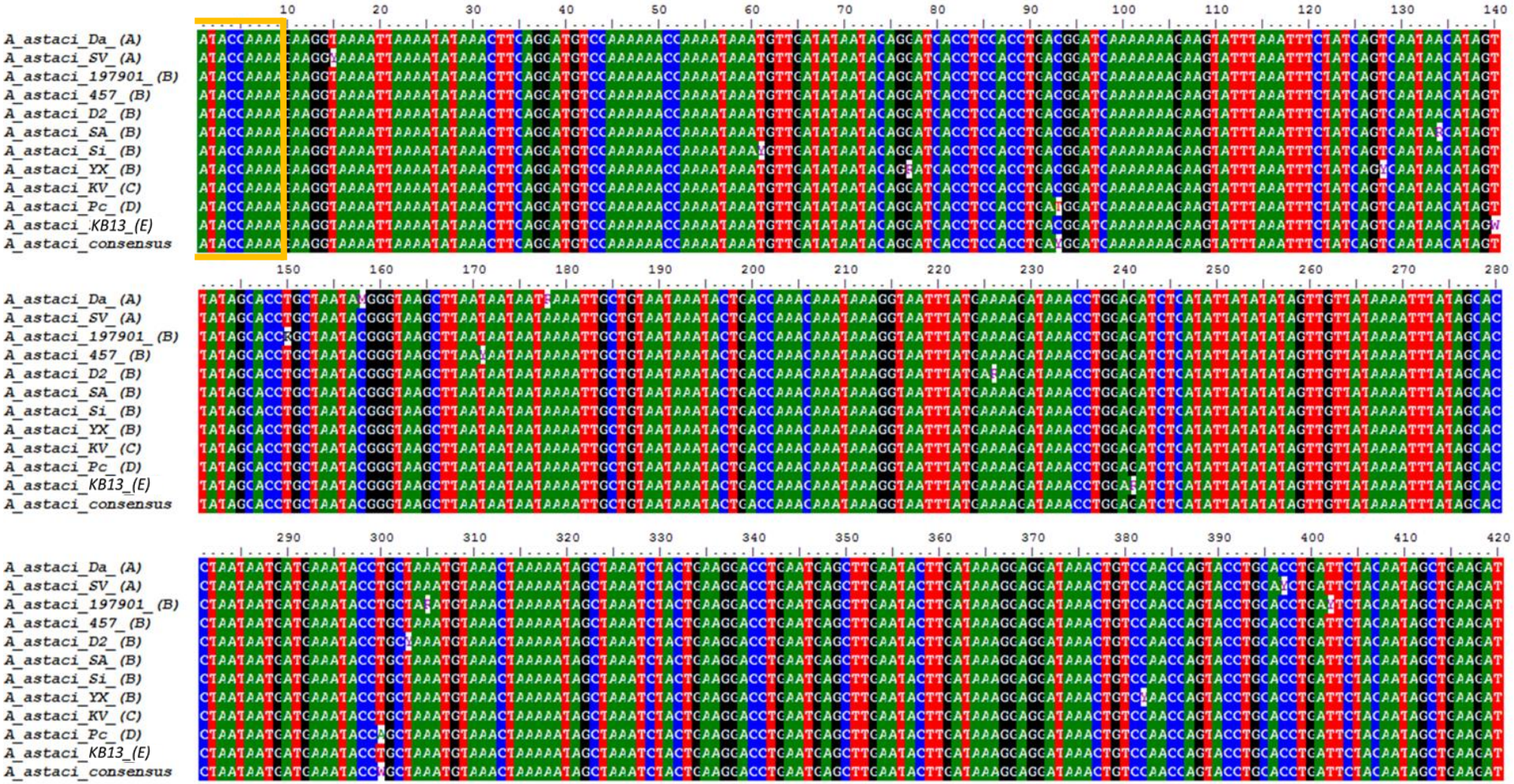
Table 5.9 List of differences in LSU cloned sequences and mutation sites. For *A. astaci* Da, Si, SV and YX and *A. invadans* NJM0002 and GWR no base changes were detected in the cloned sequences.

Species-isolate ID-clone number	Differences between clones (nt)	Variable site consensus>mutation
A. astaci 197901-21	1	845T>A
A. astaci 197901-37	1	469T>A
A. astaci 197901-6	2	1159T>A, 1166T>G
A. astaci 457-11	1	469T>A
A. astaci 457-13	1	469T>A
A. astaci 457-16	1	1388T>G
A. astaci D2-10	3	526T>C, 598T>A, 661T>C
A. astaci KV-2	4	940T>G, 944T>G, 945G>T, 946T>G
A. astaci KB13-4	7	342C>T, 491T>C, 565T>C, 1642G>A, 1814T>A, 2067G>A, 2097G>A
A. astaci Pc-2	2	93T>C, 897G>A
A. astaci SA-5	2	441C>T, 838C>T
A. frigidophilus AP5-10	1	1436C>A
A. frigidophilus RP1-18	1	1844G>A
A. frigidophilus RP2-9	1	380T>C
A. frigidophilus RP2-91	1	1338T>C
A. invadans NJM8997	4	809A>G, 820G>A, 1088T>C, 1274G>T
A. invadans NJM9030-1	2	306A>G, 345T>C
A. invadans NJM9030-8	4	349T>C, 1140A>G, 1958T>C, 2084A>G
A. invadans NJM9701-8	1	1576A>C
A. invadans-like NJM9510-8	1	45A>gap

5.5 Section 3.2.5

Table 5.10 List of differences in COI cloned sequences and mutation sites. For *A. astaci* KV and *A. frigidophilus* AP5 no base changes were detected in the cloned sequences.

Species-isolate ID-clone number	Differences between clones (nt)	Variable site consensus>mutation
A. astaci 197901-1	1	602A>C
A. astaci 197901-2	1	447T>C
A. astaci 197901-3	1	350A>G
A. astaci 197901-6	1	653T>gap
A. astaci 457-4	1	581A>G
A. astaci D2-1	1	526T>A
A. astaci D2-7	1	449A>G
A. astaci D2-10	1	227T>G
A. astaci Da-5	1	574T>C
A. astaci Da-11	1	296T>C
A. astaci Da-19	1	594G>T
A. astaci KB13-12	2	140T>A, 588A>G
A. astaci KB13-15	2	241G>A, 507 A>G
A. astaci Pc-6	1	197T>gap
A. astaci SA-1	2	38T>C, 618T>C
A. astaci Si-3	1	691A>G
A. astaci Si-6	1	76G>A
A. astaci SV-3	2	296T>G, 737A>G
A. astaci SV-5	1	355G>A
A. astaci YX-5	3	371G>A, 625A>G, 676C>T
A. frigidophilus RP1-1	1	268T>C
A. frigidophilus RP1-7	1	431T>C
A. frigidophilus RP1-8	1	64C>T
A. frigidophilus RP2-3	1	626T>C
A. frigidophilus RP2-7	1	415C>T
A. invadans GWR-6	1	312C>A
A. invadans NJM0002-2	3	122T>C, 580T>C, 713T>C
A. invadans NJM0002-7	1	198A>G
A. invadans NJM0002-10	1	449A>G
A. invadans NJM0002-17	1	728T>G
A. invadans NJM8997-6	1	359T>C
A. invadans NJM8997-9	2	35C>T, 268A>G
A. invadans NJM9030-4	2	41C>T, 72G>A
A. invadans NJM9701-5	1	458T>C
A. invadans-like NJM9510-4	1	107A>G


A_astaci_Da_(A)
A_astaci_SV_(A)
A_astaci_197901_(B)
A_astaci_457_(B)
A_astaci_D2_(B)
A_astaci_SA_(B)
A_astaci_Si_(B)
A_astaci_YX_(B)
A_astaci_KV_(C)
A_astaci_Pc_(D)
A_astaci_KB13_(E)
A_astaci_consensus

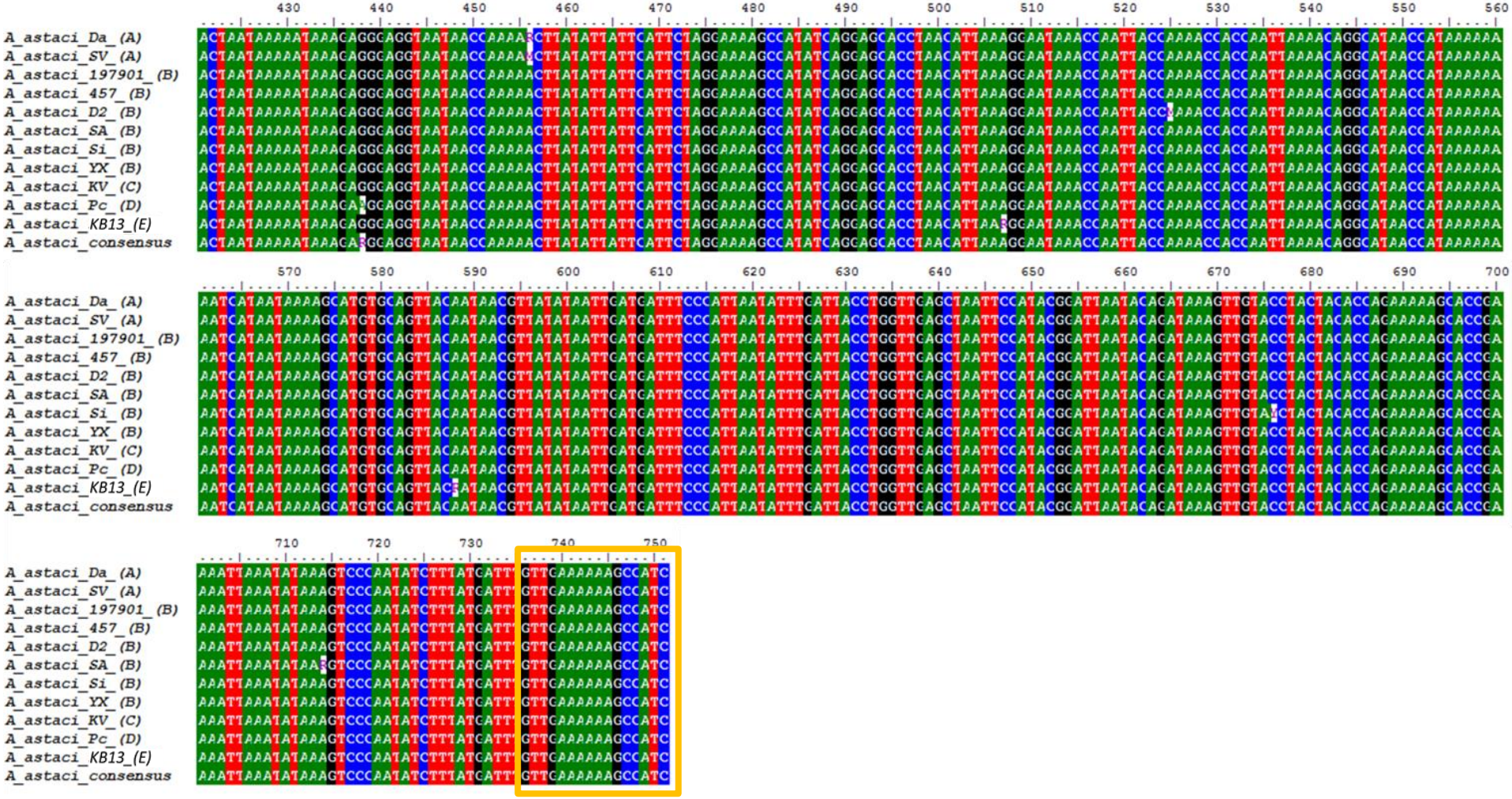


Figure 5.2 *A. astaci* COI consensus sequences showing variable sites. Consensus sequences were generated with BioEdit: for each isolate, sequences were obtained combing the respective clones; "*A. astaci* consensus" sequence was generated by combing all isolates. Variable sites for genotype D are at position 93, 330 and 438 bp. Isolate genotype indicated in brackets. Yellow boxes: primers sites.

5.6 Section 3.3.3

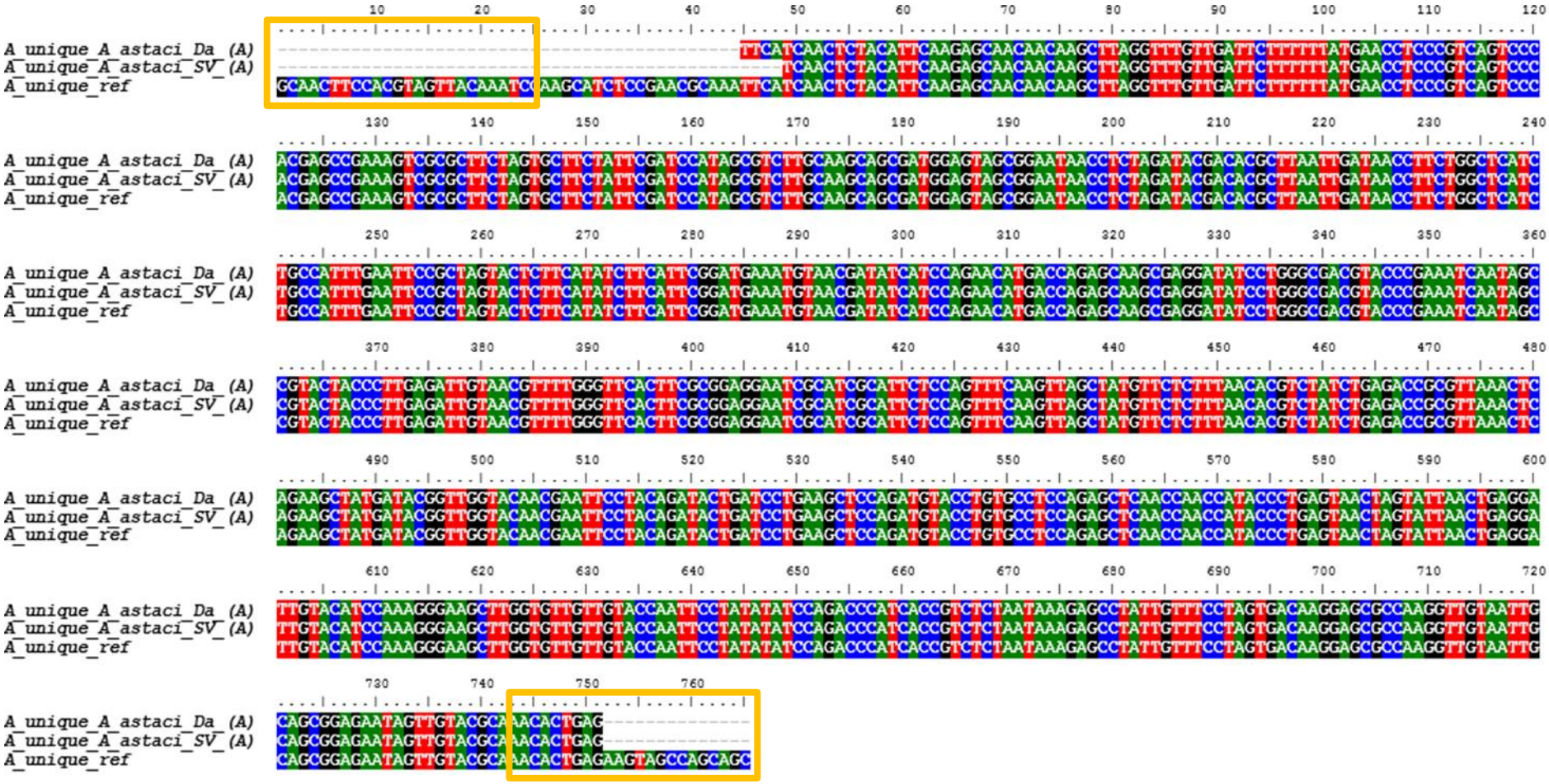


Figure 5.3 Sequences of *A. astaci* Da and SV (genotype A) amplified with A_unique primer pair. The sequences match *A. astaci* reference sequence (genotype A). Yellow boxes: primers sites.


B_unique_A_astaci_D2_(B)
B_unique_A_astaci_197901_(B)
B_unique_A_astaci_457_(B)
B_unique_A_astaci_SA_(B)
B_unique_A_astaci_Si_(B)
B_unique_A_astaci_Yx_(B)
B_unique_ref

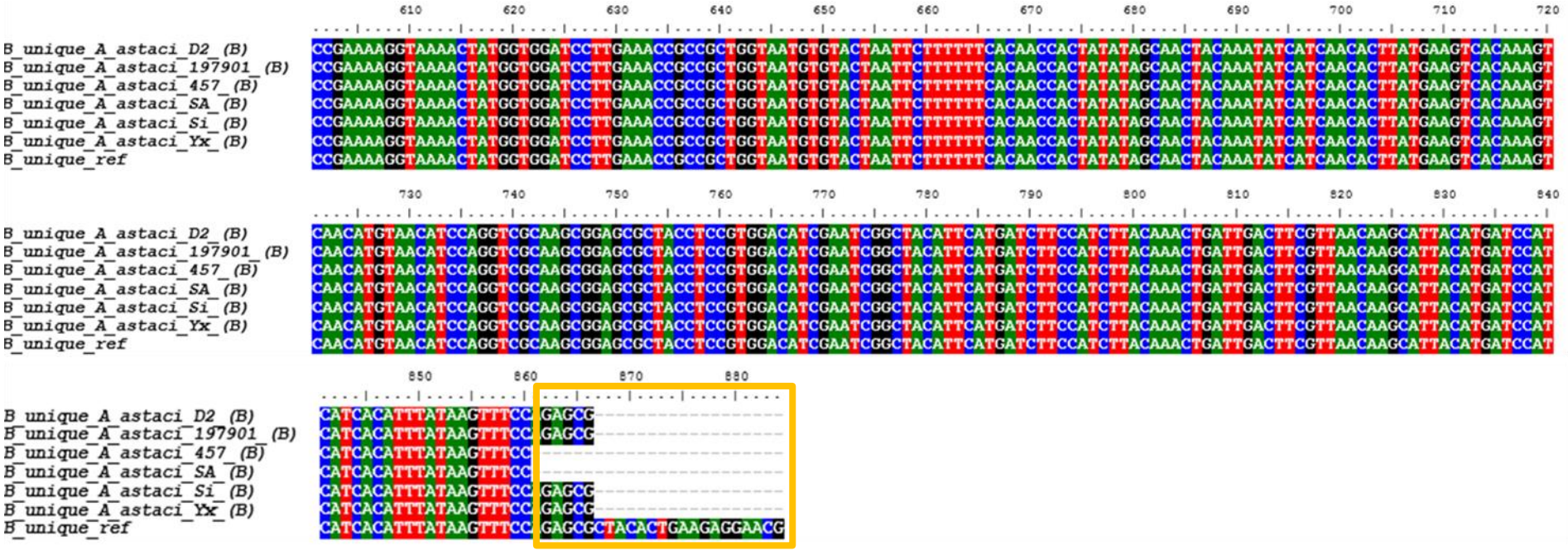


Figure 5.4 Sequences of *A. astaci* D2, 197901, 457, SA, Si, and YX (genotype B) amplified with B_unique primer pair. The sequences match the reference sequence (genotype B). Yellow boxes: primers sites.

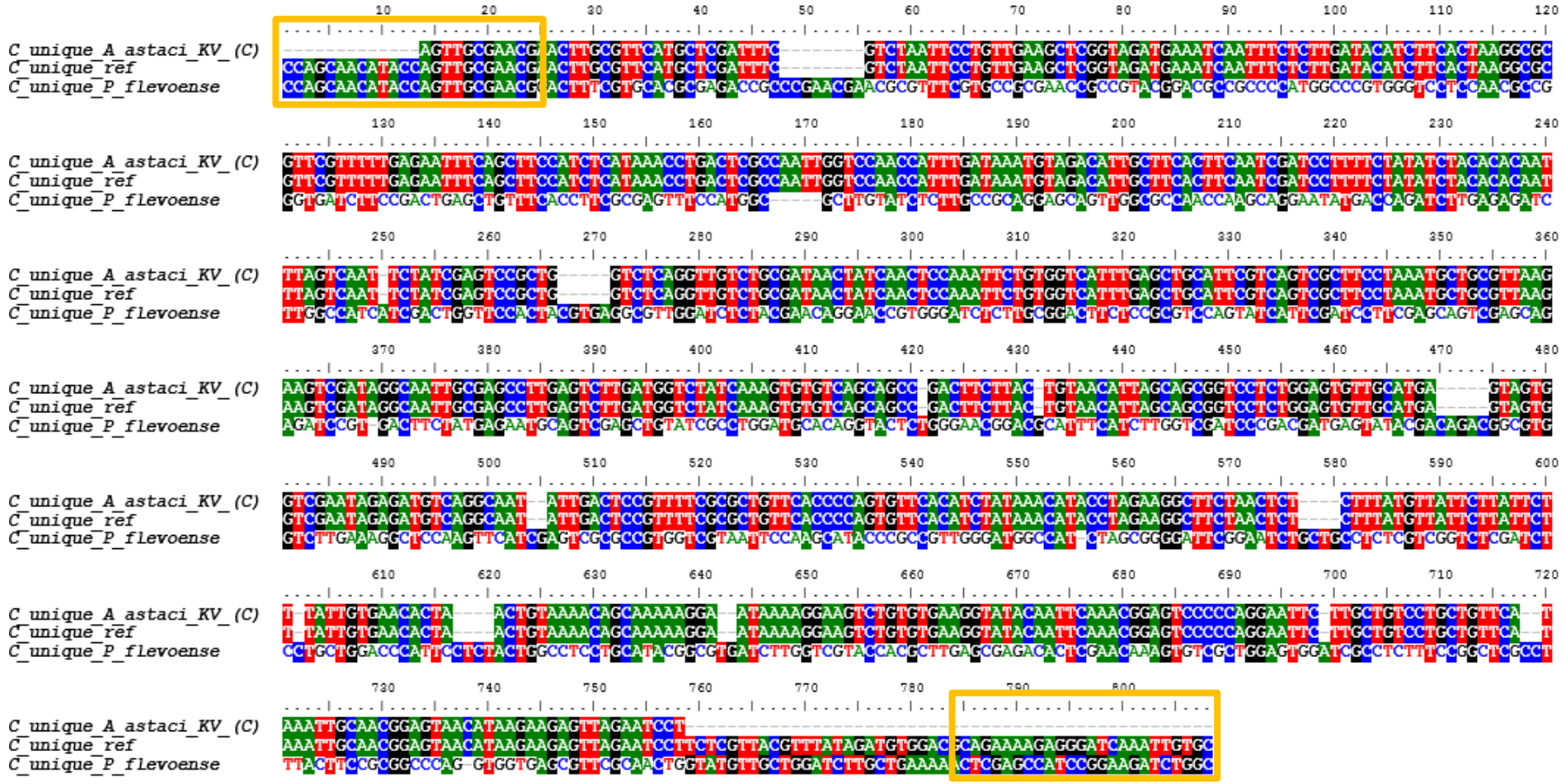


Figure 5.5 Sequences of *A. astaci* KV (genotype C) and *P. flevoense* amplified with C_unique primer pair. The sequence of *A. astaci* KV match the reference sequence (genotype C), while *P. flevoense* shares only the forward primer site. Yellow box: primers sites.

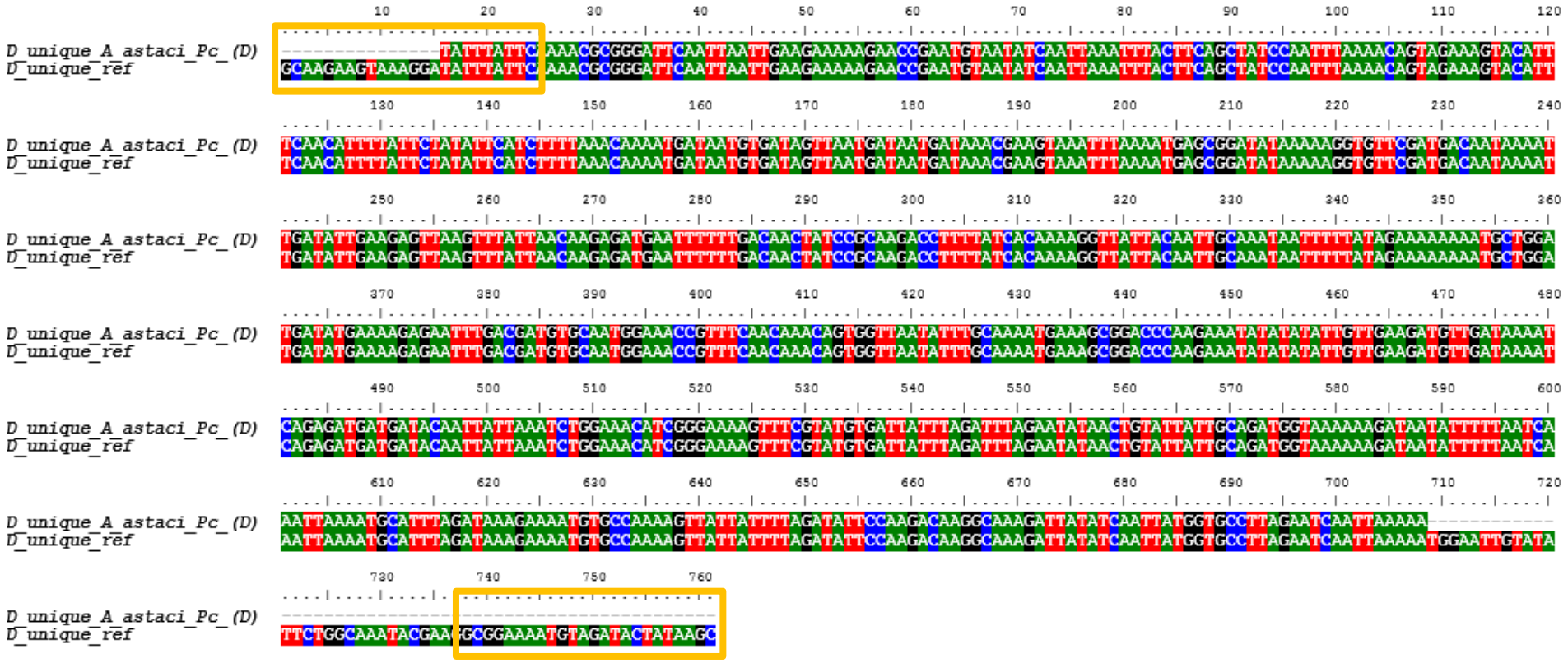


Figure 5.6 Sequences of *A. astaci* Pc (genotypeD) amplified with D_unique primer pair. The sequence matches the reference sequence (genotype D). Yellow boxes: primers sites.

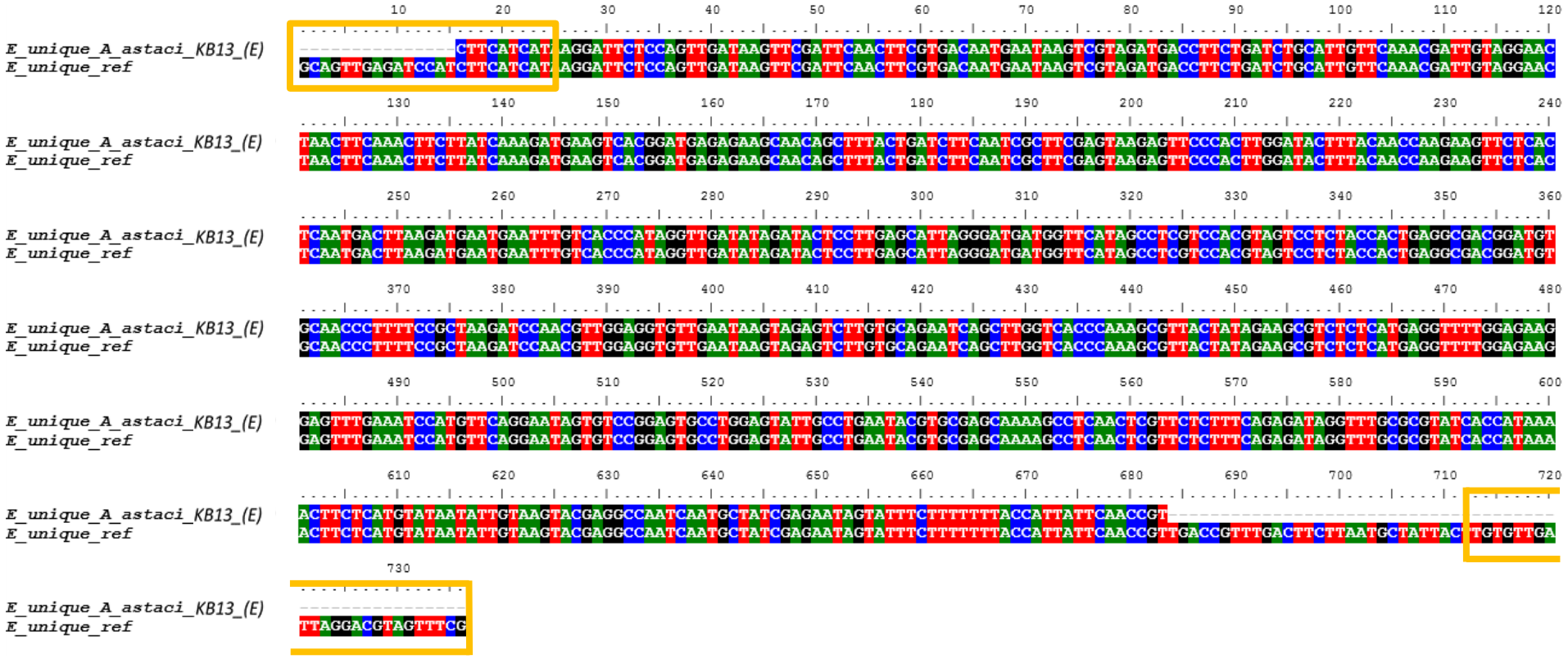


Figure 5.7 Sequences of *A. astaci* KB13 (genotype E) amplified with E_unique primer pair. The sequence match the reference sequence (genotype E). Yellow boxes: primers sites.

5.7 Section 3.4.3

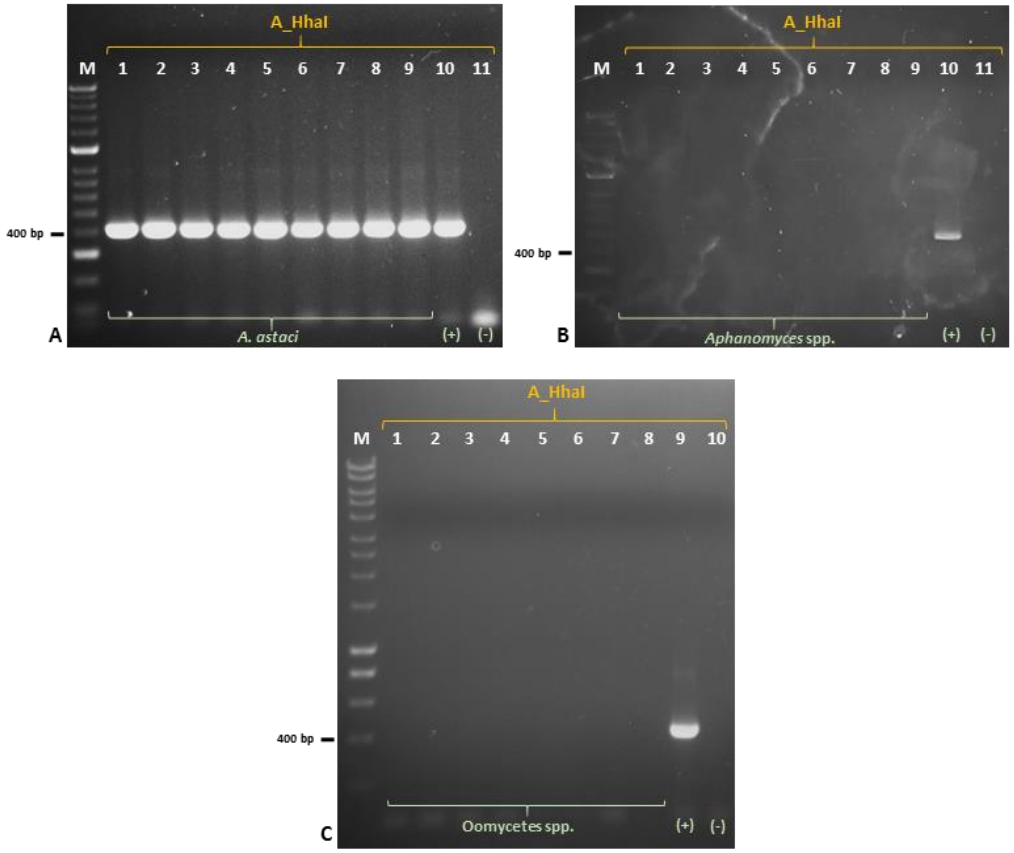


Figure 5.8 PCR on *Aphanomyces* and oomycetes isolates with A_HhaI primer pair. **A**: M, Bioline HyperLadder II; 1, *A. astaci* 457 (genotype B); 2, *A. astaci* 197901 (genotype B); 3, *A. astaci* SA (genotype B); 4, *A. astaci* YX (genotype B); 5, *A. astaci* D2 (genotype B); 6, *A. astaci* Si (genotype B); 7, *A. astaci* KV (genotype C); 8, *A. astaci* Pc (genotype D); 9, *A. astaci* KB13 (genotype E); 10, *A. astaci* Da (genotype A); 11, negative control. **B**: M, Bioline HyperLadder II; 1, *A. invadans* NJM9701; 2, *A. invadans* NJM8997; 3, *A. invadans* NJM9030; 4, *A. invadans* NJM0002; 5, *A. invadans* GRW; 6, *A. invadans*-like NJM9510; 7, *A. frigidophilus* AP5; 8, *A. frigidophilus* RP1; 9, *A. frigidophilus* RP2; 10, *A. astaci* Da (genotype A); 11, negative control. **C**: M, Bioline HyperLadder I; 1, *S. parasitica*; 2, *S. furcata*; 3, *L. caudata*; 4, *A. racemosa*; 5, *A. laevis*; 6, *P. monospermum*; 7, *P. flevoense*; 8, *Phoma*-like; 9, *A. astaci* Da (genotype A); 10, negative control. (+), positive control; (-), negative control.

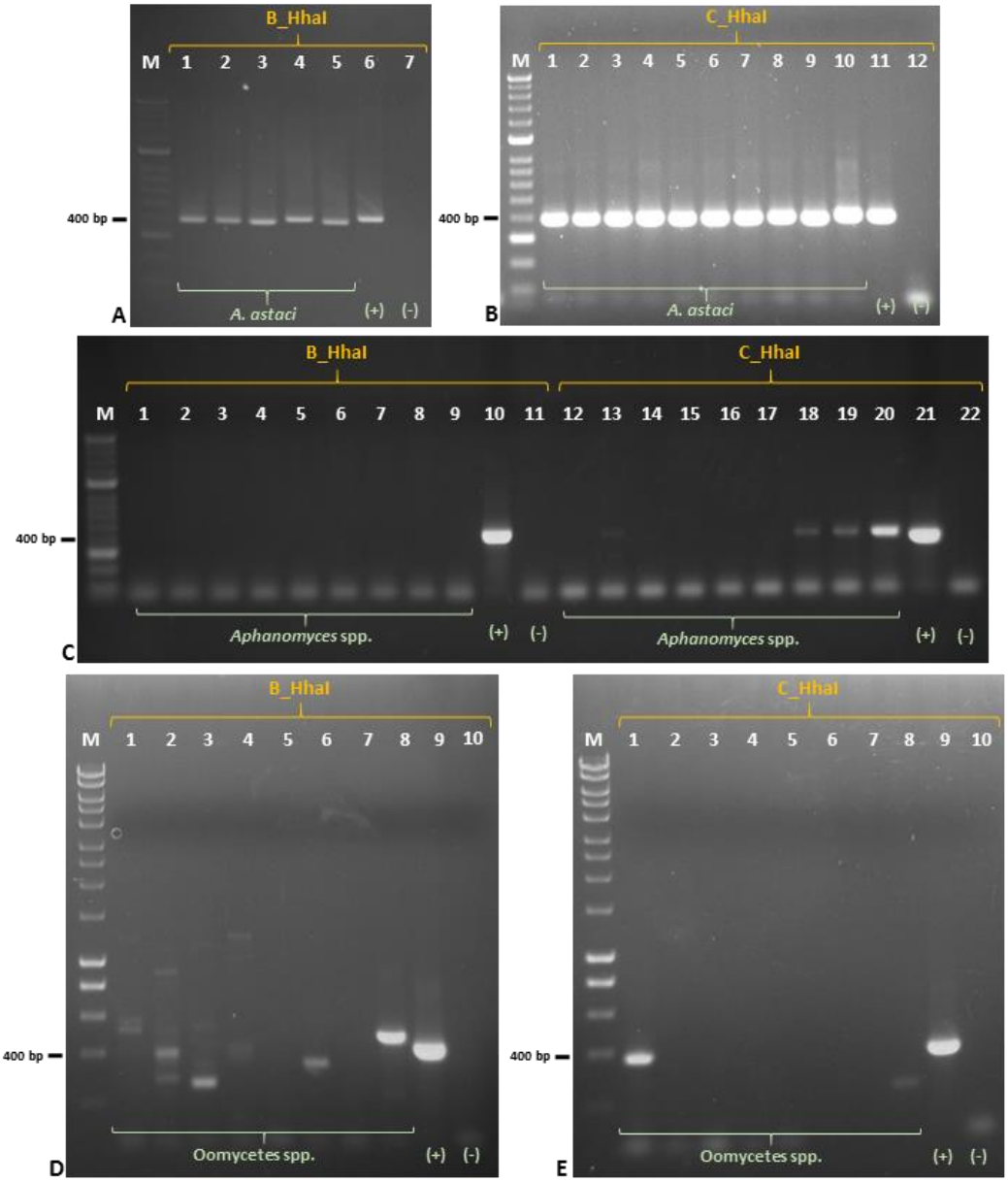


Figure 5.9 PCR on *Aphanomyces* and oomycetes isolates with B_HhaI and C_HhaI primer pairs. **A**: M, Bioline HyperLadder II; 1, *A. astaci* SV (genotype A); 2, *A. astaci* Da (genotype A); 3, *A. astaci* KV (genotype C); 4, *A. astaci* Pc (genotype D); 5, *A. astaci* KB13 (genotype E); 6, *A. astaci* D2 (genotype B); 7, negative control. **B**: M, Bioline HyperLadder II; 1, *A. astaci* SV (genotype A); 2, *A. astaci* Da (genotype A); 3, *A. astaci* 457 (genotype B); 4, *A. astaci* 197901 (genotype B); 5, *A. astaci* SA (genotype B); 6, *A. astaci* YX (genotype B); 7, *A. astaci* D2 (genotype B); 8, *A. astaci* Si (genotype B); 9, *A. astaci* Pc (genotype D); 10, *A. astaci* KB13 (genotype E); 11, *A. astaci* KV (genotype C); 12, negative control. **C**: M, Bioline HyperLadder II; 1, *A. invadans* NJM9701; 2, *A. invadans* NJM8997; 3, *A. invadans* NJM9030; 4, *A. invadans* NJM0002; 5, *A. invadans* GRW; 6, *A. invadans*-like NJM9510; 7, *A. frigidophilus* RP1; 8, *A. frigidophilus* RP2; 9, *A. frigidophilus* AP5; 10, *A. astaci* D2 (genotype B); 11, negative control; 12, *A. invadans* NJM9701; 13, *A. invadans* NJM8997; 14, *A. invadans* NJM9030; 15, *A. invadans* NJM0002; 16, *A. invadans* GRW; 17, *A. invadans*-like NJM9510; 18, *A. frigidophilus* RP1; 19, *A. frigidophilus* RP2; 20, *A. frigidophilus* AP5; 21, *A. astaci* KV (genotype C); 22, negative control. **D** and **E**: M, Bioline HyperLadder I; 1, *S. parasitica*; 2, *S. furcata*; 3, *L. caudata*; 4, *A. racemosa*; 5, *A. laevis*; 6, *P. monospermum*; 7, *P. flevoense*; 8, *Phoma*-like; 9 **D**, *A. astaci* D2 (genotype B); 9 **E**, *A. astaci* KV (genotype C); 10, negative control. (+), positive control; (-), negative control.

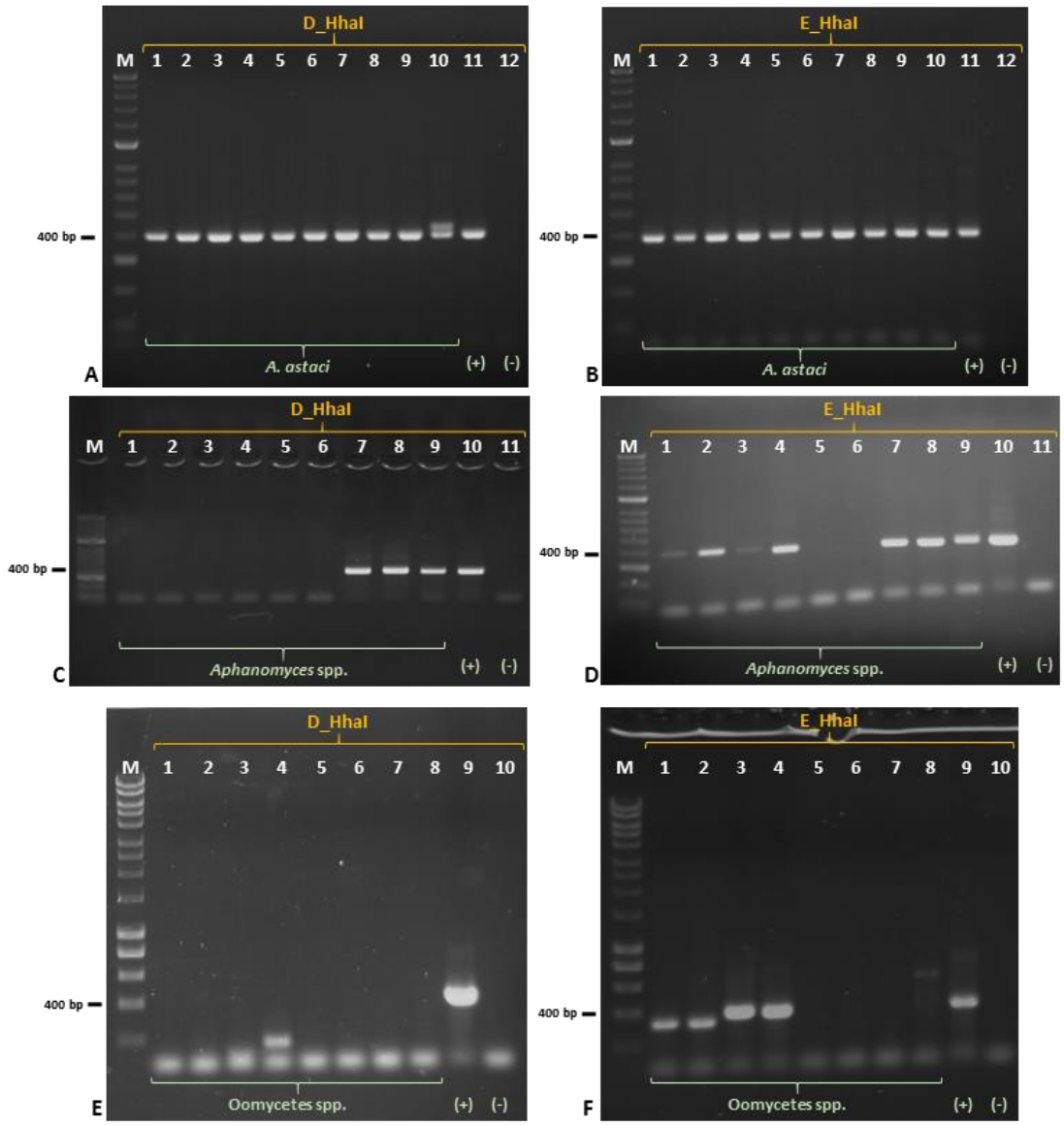


Figure 5.10 PCR on *Aphanomyces* and oomycetes isolates with D_HhaI and E_HhaI primer pairs. **A**: M, Bioline HyperLadder II; 1, *A. astaci* SV (genotype A); 2, *A. astaci* Da (genotype A); 3, *A. astaci* 457 (genotype B); 4, *A. astaci* 197901 (genotype B); 5, *A. astaci* SA (genotype B); 6, *A. astaci* YX (genotype B); 7, *A. astaci* D2 (genotype B); 8, *A. astaci* Si (genotype B); 9, *A. astaci* KV (genotype C); 10, *A. astaci* KB13 (genotype E); 11, *A. astaci* Pc (genotype D); 12, negative control. **B**: M, Bioline HyperLadder II; 1, *A. astaci* SV (genotype A); 2, *A. astaci* Da (genotype A); 3, *A. astaci* 457 (genotype B); 4, *A. astaci* 197901 (genotype B); 5, *A. astaci* SA (genotype B); 6, *A. astaci* YX (genotype B); 7, *A. astaci* D2 (genotype B); 8, *A. astaci* Si (genotype B); 9, *A. astaci* KV (genotype C); 10, *A. astaci* Pc (genotype D); 11, *A. astaci* KB13 (genotype E); 12, negative control. **C** and **D**: M, Bioline HyperLadder II; 1, *A. invadans* NJM9701; 2, *A. invadans* NJM8997; 3, *A. invadans* NJM9030; 4, *A. invadans* NJM0002; 5, *A. invadans* GRW; 6, *A. invadans*-like NJM9510; 7, *A. frigidophilus* RP1; 8, *A. frigidophilus* RP2; 9, *A. frigidophilus* AP5; 10, **C**, *A. astaci* Pc (genotype D); 10 **D**, *A. astaci* KB13 (genotype E); 11, negative control. **E** and **F**: M, Bioline HyperLadder I; 1, *S. parasitica*; 2, *S. furcata*; 3, *L. caudata*; 4, *A. racemosa*; 5, *A. laevis*; 6, *P. monospermum*; 7, *P. flevoense*; 8, *Phoma*-like; 9 **E**, *A. astaci* Pc (genotype D); 9 **F**, *A. astaci* KB13 (genotype E); 10, negative control. (+), positive control; (-), negative control.

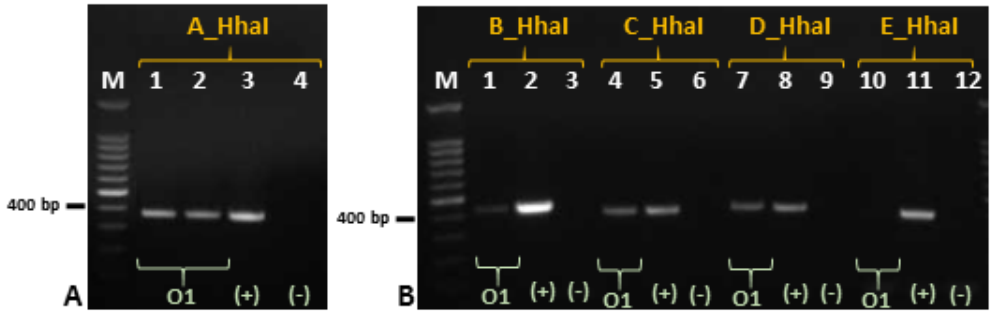


Figure 5.11 PCR on outbreak PM28325 (O1). **A**: 1, PM28325 1.2; 2, PM28325 1.4; 3, *A. astaci* Da (genotype A); 4, negative control. **B**: 1, PM28325 1.2; 2, *A. astaci* D2 (genotype B); 3, negative control; 4, PM28325 1.2; 5, *A. astaci* KV (genotype C); 6, negative control; 7, PM28325 1.2; 8; *A. astaci* Pc (genotype D); 9 negative control; 10, PM28325 1.2; 11, *A. astaci* KB13 (genotype E); 12, negative control. M, Promega 100 bp DNA Ladder. (+), positive control; (-), negative control.

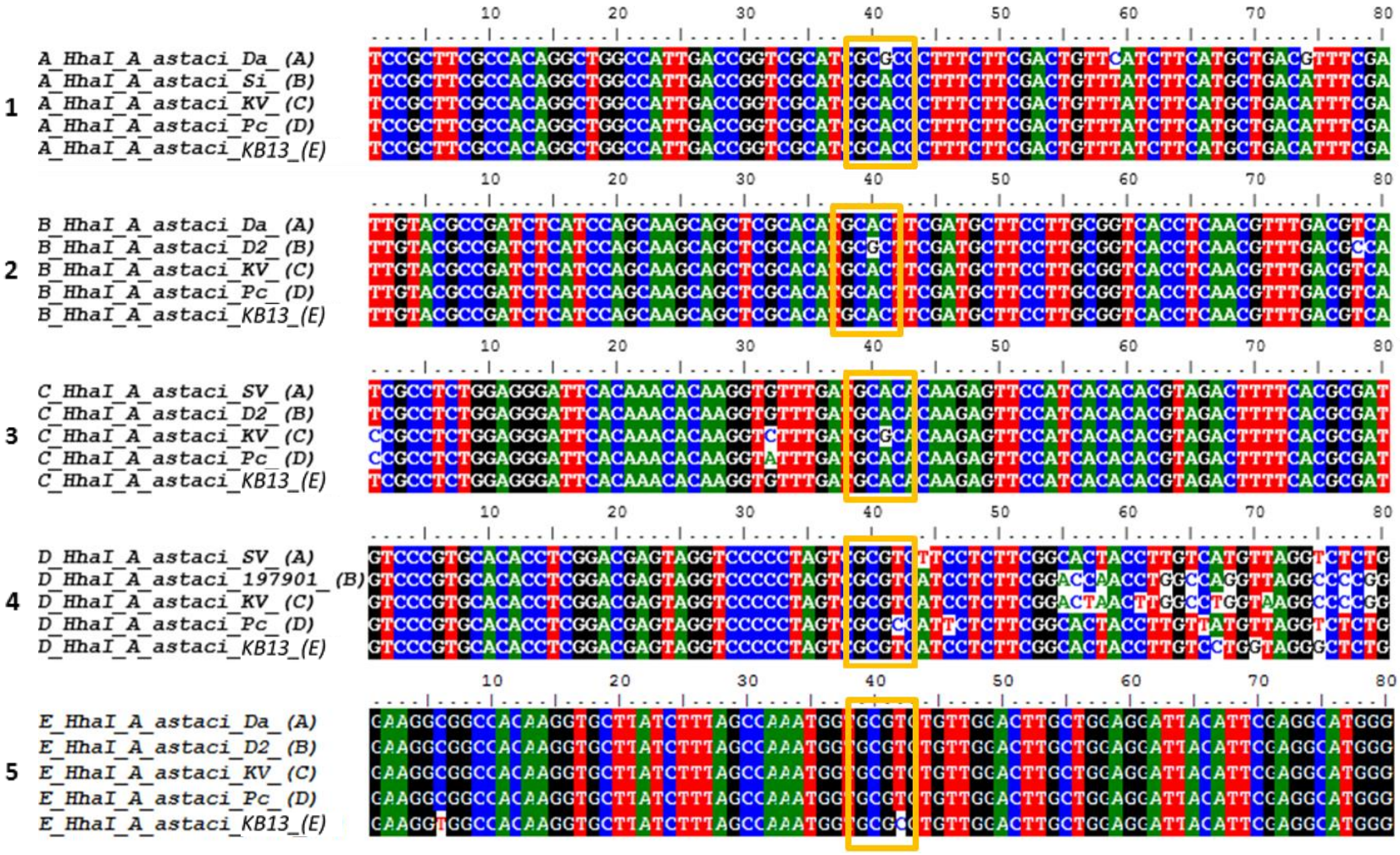


Figure 5.12 Restriction sites in nuclear DNA of *A. astaci* from sequenced PCR products. PRC products obtained with primer pairs A_HhaI (**1**), B_HhaI (**2**), C_HhaI (**3**), D_HhaI (**4**) and E_HhaI (**5**). Yellow box: restriciton site ("GCGC"). Isolates genotypes indicated in brackets.

5.8 Section 3.4.4

Table 5.11 *A. astaci* gDNA dilutions used for sensitivity tests. Column 2: concentration of gDNA in original samples measured with Qubit® 2.0 Fluorometer (Invitrogen). Columns 3 to 6 final gDNA concentration in 20 µl PCR reaction.

		Final dilutions in 20 µl PCR reaction					
Isolate	**1 ng/µl**	**10^{-1} ng/µl**	**10^{-3} pg/µl**	**10^{-5} fg/µl**	**10^{-7} fg/µl**	**10^{-9} ag/µl**	**10^{-10} ag/µl**
Da	78.3	0.39	3.9	39	0.39	3.9	0.39
D2	314	1.57	15.7	157	1.57	15.7	1.57
KV	289	1.44	14.4	144	1.44	14.4	1.44
Pc	380	1.9	19	190	1.9	19	1.9
KB13	176	0.88	8.8	88	0.88	8.8	0.88

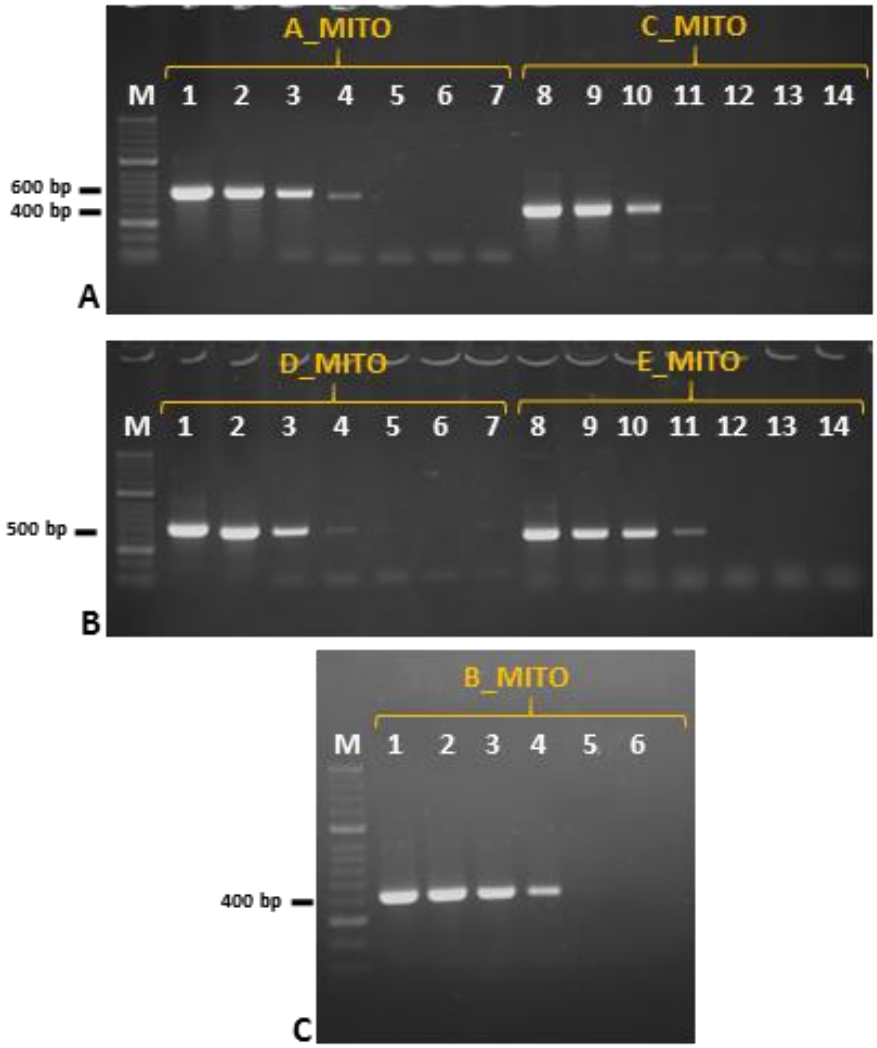


Figure 5.13 Sensitivity test of MITO primers on diluted *A. astaci* gDNA. **A**: 1-6, *A. astaci* Da amplified with A_MITO primers; 7, negative control; 8-13, *A. astaci* KV amplified with C_MITO primers; 14, negative control. **B**: 1-6, *A. astaci* Pc amplified with D_MITO primers; 7, negative control; 8-13, *A. astaci* KB13 amplified with E_MITO primers; 14, negative control. **C**: 1-5, *A. astaci* D2 amplified with B_MITO primers; 6, negative control. M: Bioline HyperLadder II. Dilutions for genotypes A, C, D and E: 10^{-1}, 10^{-3}, 10^{-5}, 10^{-7}, 10^{-9}, 10^{-10}; genotype B: 10^{-1}, 10^{-3}, 10^{-5}, 10^{-7}, 10^{-9}.

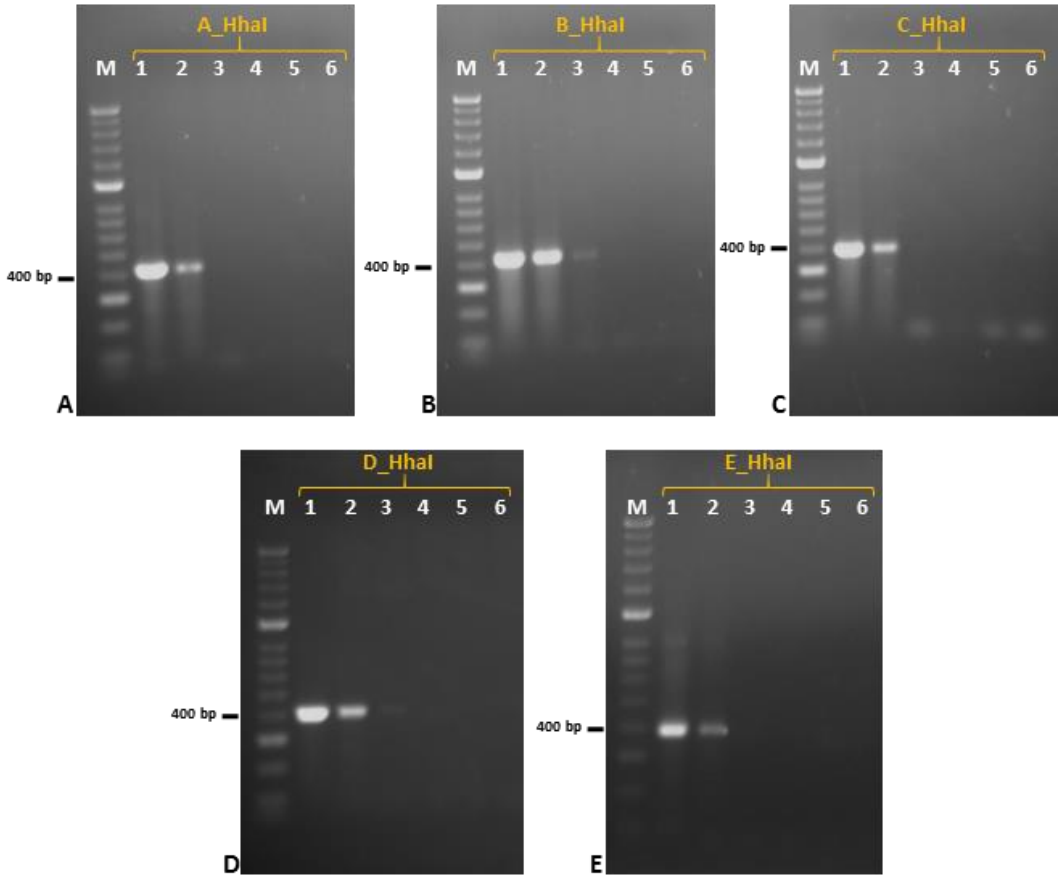


Figure 5.14 Sensitivity test of primers targeting the nuclear DNA on diluted *A. astaci* gDNA. **A**: 1-5, *A. astaci* Da amplified with A_HhaI primers; 6, negative control; **B**: 1-5, *A. astaci* D2 amplified with B_HhaI primers; 6, negative control; **C**: 1-5, *A. astaci* KV amplified with C_HhaI primers; 6, negative control; **D**: 1-5, *A. astaci* Pc amplified with D_HhaI primers; 6, negative control; **E**: 1-5, *A. astaci* KB13 amplified with E_HhaI primers; 6, negative control. M: Bioline HyperLadder II. Dilutions: 10^{-1}, 10^{-3}, 10^{-5}, 10^{-7}, 10^{-9}.

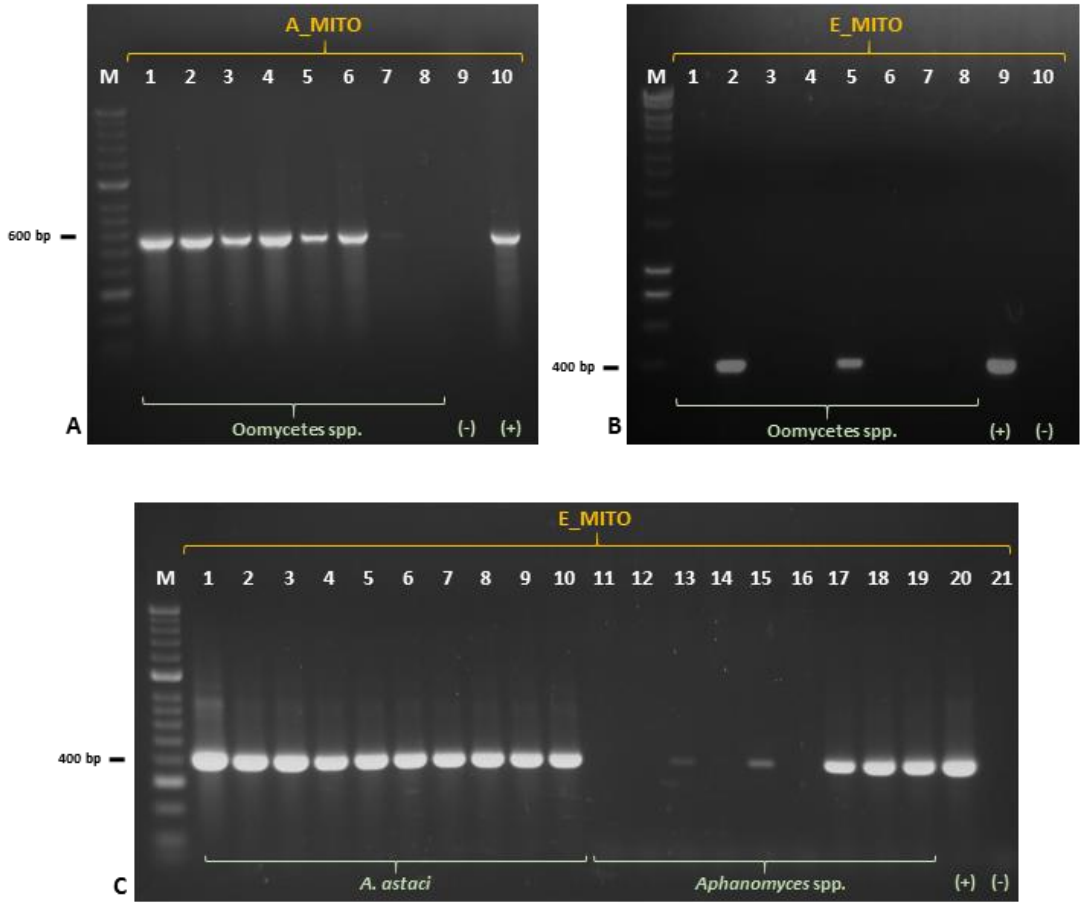


Figure 5.15 PCR on oomycetes isolates with A_MITO (**A**) and E_MITO (**B** and **C**) primer pairs. **A**: M, Bioline HyperLadder II; 1, *S. parasitica*; 2, *S. furcata*; 3, *L. caudata*; 4, *A. racemosa*; 5, *A. laevis*; 6, *P. monospermum*; 7, *P. flevoense*; 8, *Phoma*-like; 9, negative control; 10, *A. astaci* Da (genotype A). **B**: M, Bioline HyperLadder I; 1, *S. parasitica*; 2, *S. furcata*; 3, *L. caudata*; 4, *A. racemosa*; 5, *A. laevis*; 6, *P. monospermum*; 7, *P. flevoense*; 8, *Phoma*-like; 9, *A. astaci* KB13 (genotype E); 10, negative control. **C**: M, Bioline HyperLadder II; 1, *A. astaci* SV (genotype A); 2, *A. astaci* Da (genotype A); 3, *A. astaci* D2 (genotype B); 4, *A. astaci* YX (genotype B); 5, *A. astaci* 457 (genotype B); 6, *A. astaci* 197901 (genotype B); 7, *A. astaci* Si (genotype B); 8, *A. astaci* SA (genotype B); 9, *A. astaci* KV (genotype C); 10, *A. astaci* Pc (genotype D); 11, *A. invadans* NJM8997; 12, *A. invadans* NJM9030; 13, *A. invadans* NJM9701; 14, *A. invadans* NJM0002; 15, *A. invadans* GRW; 16, *A. invadans*-like NJM9510; 17, *A. frigidophilus* AP5; 18, *A. frigidophilus* RP1; 19, *A. frigidophilus* RP2; 20, *A. astaci* KB13 (genotype E); 21, negative control. (+), positive control; (-), negative control.

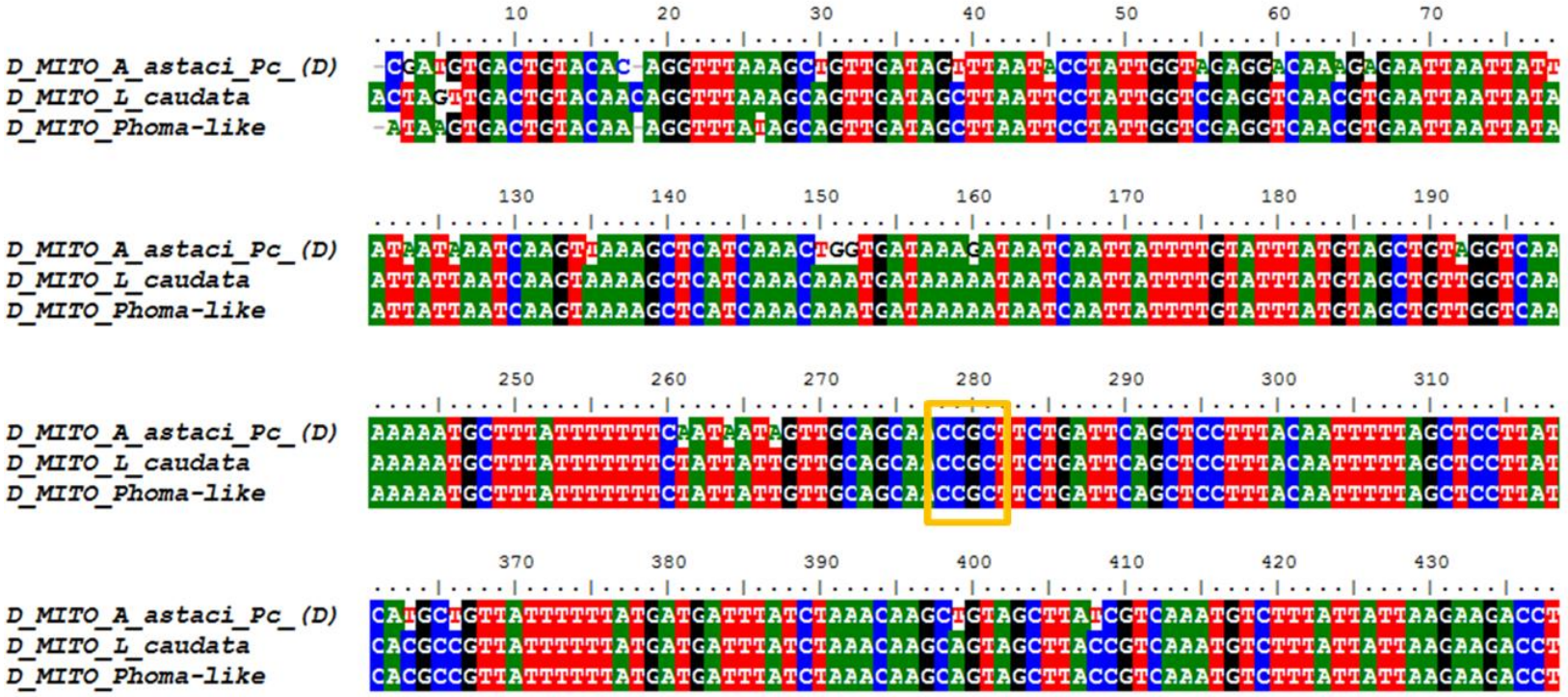


Figure 5.16 Sequences of PCR products amplified with D_MITO primers. *L. caudata* and *Phoma*-like shown similar digestion patterns to *A. astaci* Pc (genotype D). Aligning their purified PCR products revealed the presence of the same SNV in the restriction site for A. astaci Pc (genotype D), but overall sequence difference between the species. Yellow box: SNV in restriction site.

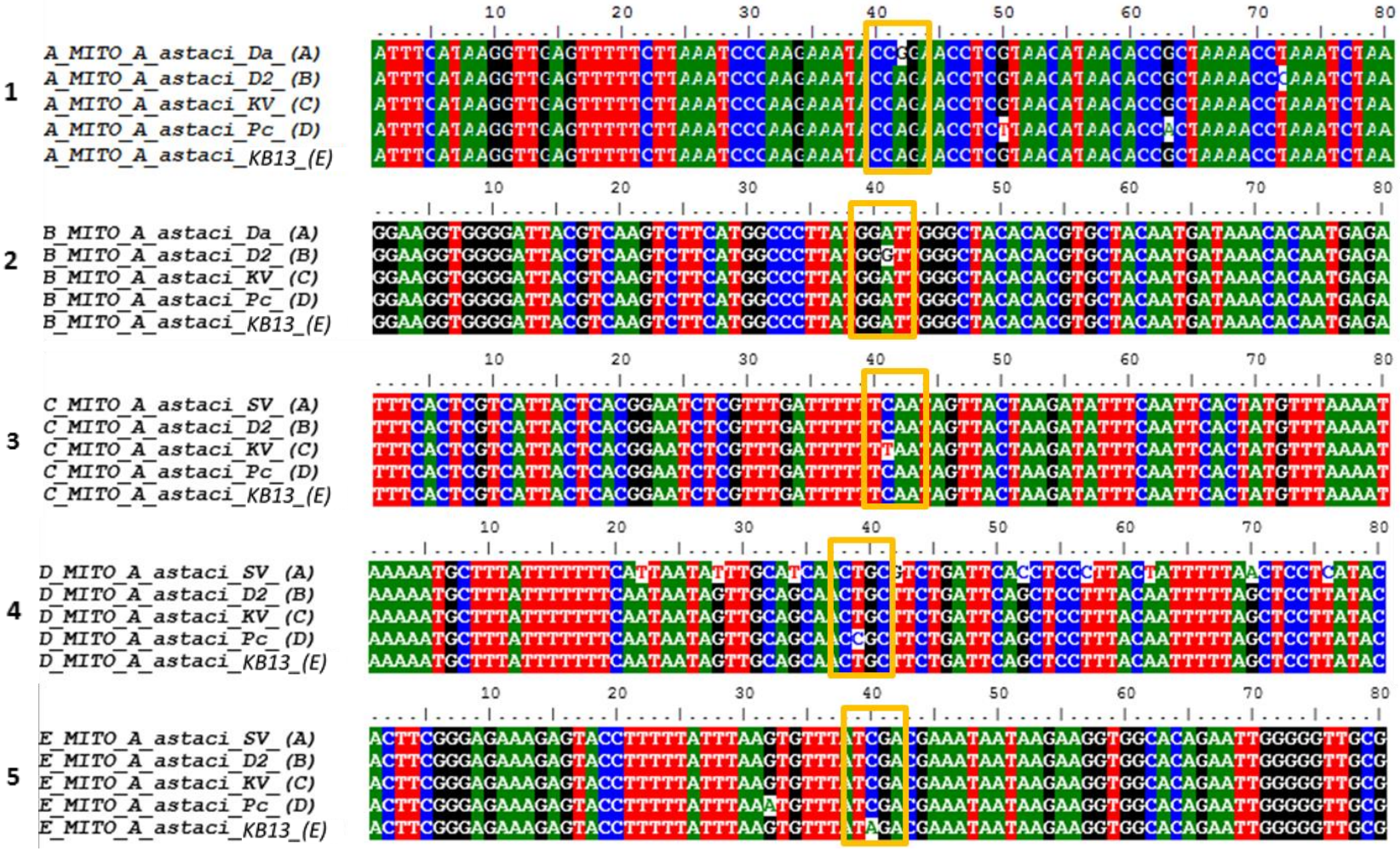


Figure 5.17 *A. astaci* SNVs in mtDNA from sequenced PCR products. PRC products obtained with primer pairs A_MITO (**1**), B_MITO (**2**), C_MITO (**3**), D_MITO (**4**) and E_MITO (**5**). Yellow box in alignement number 1 and 3-5: SNVs in restriciton sites. Yellow box in alignement number 2: SNV selected for the development of the semi'nested PCR assay. Isolates genotypes indicated in brackets.

5.9 Section 3.4.6

Table 5.12 PCR results from outbreaks samples amplified by A, B, C, D, and E mitochondrial primer pairs. +/orange: sample positive for PCR reaction; -/light blue: sample negative for PCR reaction.

		PCR 45 cycles with mitochondrial primer pairs on outbreaks samples									
Outbreak	**Sample**	**A_MITO**		**B_MITO**		**C_MITO**		**D_MITO**		**E_MITO**	
PM28325	**1.2**	+	Figure 5.18 A well 1	+	Figure 5.18 A well 6	+	Figure 5.18 B well 1	+	Figure 5.18 B well 6	+	Figure 5.18 B well 11
	1.4	+	Figure 5.18 A well 2	+	Figure 5.18 A well 7	+	Figure 5.18 B well 2	+	Figure 5.18 B well 7	+	Figure 5.18 B well 12
PM28465	**1.2**	-	Figure 5.18 A well 3	+	Figure 5.18 A well 8	+	Figure 5.18 B well 3	-	Figure 5.18 B well 8	+	Figure 5.18 B well 13
PM21018	**1**	-	Figure 5.19 A well 1	+	Figure 5.19 B Well 1	+	Figure 5.19 C Well 1	+	Figure 5.19 D Well 1	+	Figure 5.19 E Well 1
	2	+	Figure 5.19 A Well 2	-	Figure 5.19 B Well 2	+	Figure 5.19 C Well 2	-	Figure 5.19 D Well 2	-	Figure 5.19 E Well 2
	3	-	Figure 5.19 A Well 3	+	Figure 5.19 B Well 3	-	Figure 5.19 C Well 3	-	Figure 5.19 D Well 3	+	Figure 5.19 E Well 3
PM19790	**1**	+	Figure 5.19 A Well 4	+	Figure 5.19 B Well 4	+	Figure 5.19 C Well 4	+	Figure 5.19 D Well 4	+	Figure 5.19 E Well 4
	2	+	Figure 5.19 A Well 5	+	Figure 5.19 B Well 5	+	Figure 5.19 C Well 5	+	Figure 5.19 D Well 5	+	Figure 5.19 E Well 5
	6	+	Figure 5.19 A Well 6	+	Figure 5.19 B Well 6	+	Figure 5.19 C Well 6	+	Figure 5.19 D Well 6	+	Figure 5.19 E Well 6
	7	+	Figure 5.19 A Well 7	+	Figure 5.19 B Well 7	+	Figure 5.19 C Well 7	+	Figure 5.19 D Well 7	+	Figure 5.19 E Well 7
PM19955	**1.1**	+	Figure 5.19 A Well 8	+	Figure 5.19 B Well 8	+	Figure 5.19 C Well 8	+	Figure 5.19 D Well 8	+	Figure 5.19 E Well 8
	2.1	-	Figure 5.19 A Well 9	-	Figure 5.19 B Well 9	+	Figure 5.19 C Well 9	-	Figure 5.19 D Well 9	-	Figure 5.19 E Well 9
	3.1	+	Figure 5.19A Well 10	+	Figure 5.19 B Well 10	+	Figure 5.19 C Well 10	+	Figure 5.19 D Well 10	+	Figure 5.19 E Well 10
PM-M17120	**1**	-	Figure 5.19 A Well 11	-	Figure 5.19 B Well 11	-	Figure 5.19 C Well 11	-	Figure 5.19 D Well 11	-	Figure 5.19 E Well 11
	2	-	Figure 5.19 A Well 12	-	Figure 5.19 B Well 12	-	Figure 5.19 C Well 12	-	Figure 5.19 D Well 12	-	Figure 5.19 E Well 12

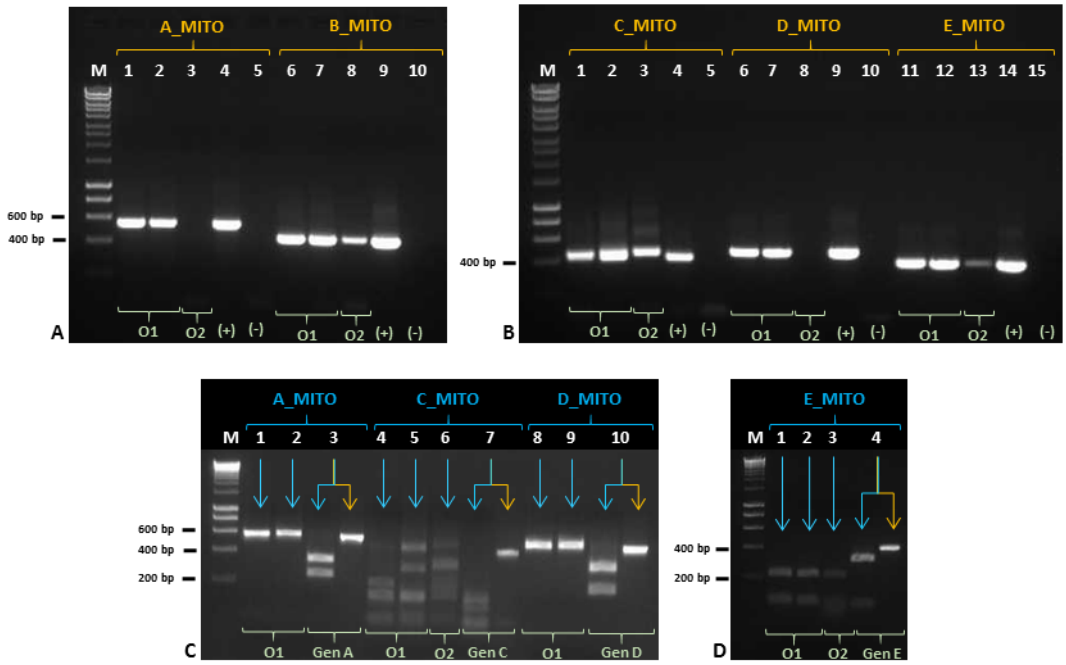


Figure 5.18 Outbreak samples O1 and O2 PCR (**A** and **B**) and enzymatic digestions (**C** and **D**) results with mitochondrial primer pairs. **A**, A_MITO and B_MITO primer pairs: 1, O1-1.2; 2, O1-1.4; 3, O2-1.2; 4, *A. astaci* Da (genotype A); 5, negative control; 6, O1-1.2; 7, O1-1.4; 8, O2-1.2; 9, *A. astaci* D2 (genotype B); 10, negative control. **B**, C_MITO, D_MITO and E_MITO primer pairs: 1, O1-1.2; 2, O1-1.4; 3, O2-1.2; 4, *A. astaci* KV (genotype C); 5, negative control; 6, O1-1.2; 7, O1-1.4; 8, O2-1.2; 9, *A. astaci* Pc (genotype D); 10, negative control; 11, O1-1.2; 12, O1-1.4; 13, O2-1.2; 14, *A. astaci* KB13 (genotype E); 15, negative control. **C**, A_MITO, C_MITO and D_MITO primer pairs: 1, O1-1.2; 2, O1-1.4; 3, *A. astaci* Da (genotype A); 4, O1-1.2; 5, O1-1.4; 6, O2-1.2; 7, *A. astaci* KV (genotype C); 8, O1-1.2; 9, O1-1.4; *A. astaci* Pc (genotype D). **D**, E_MITO primer pair: 1, O1-1.2; 2, O1-1.4; 3, O2-1.2; 4, *A. astaci* KB13 (genotype E). M, Bioline HyperLadder I. (+), positive control; (-), negative control. Blue arrows: PCR product after enzymatic digestion. Yellow arrows: original PCR product. For outbreaks description see section 2.1.8.

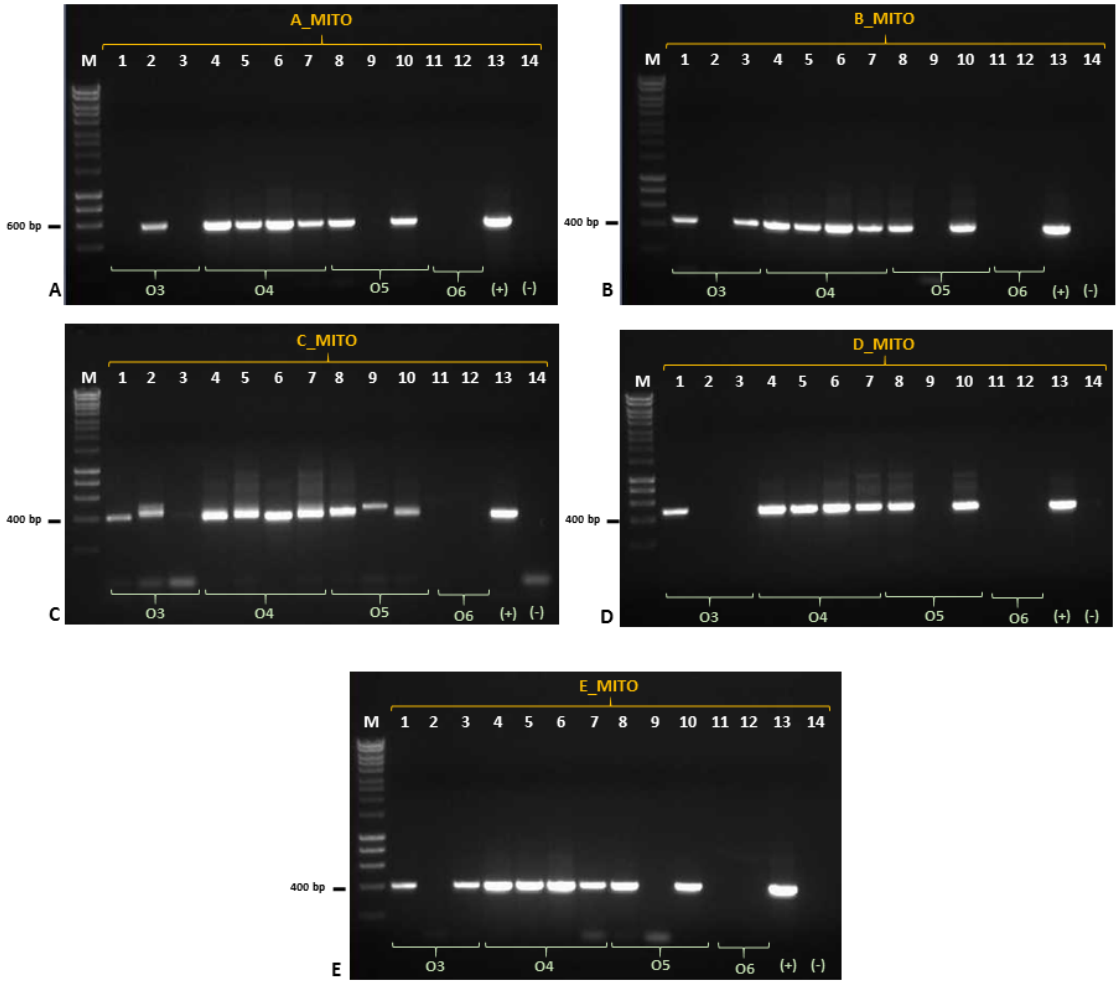


Figure 5.19 Outbreak samples O3 to O6 PCR amplification results with mitochondrial primer pairs. **A**: A_MITO primer pair; **B**: B_MITO primer pair; **C**: C_MITO primer pair; **D**: D_MITO primer pair; **E**: E_MITO primer pair. In all figures: M, Bioline HyperLadder I; 1, O3-1; 2, O3-2; 3, O3-3; 4, O4-1; 5, O4-2; 6, O4-6; 7, O4-7; 8, O5-1.1; 9, O5-2.1; 10, O5-3.1; 11, O6-1; 12, O6-2; **A13**, *A. astaci* Da (genotype A); **B13**, *A. astaci* D2 (genotype B); **C13**, *A. astaci* KV (genotype C); **D13**, *A. astaci* Pc (genotype D); **E13**, *A. astaci* KB13 (genoytpe E); 14, negative control. (+), positive control; (-), negative control. For outbreaks description see section 2.1.8.

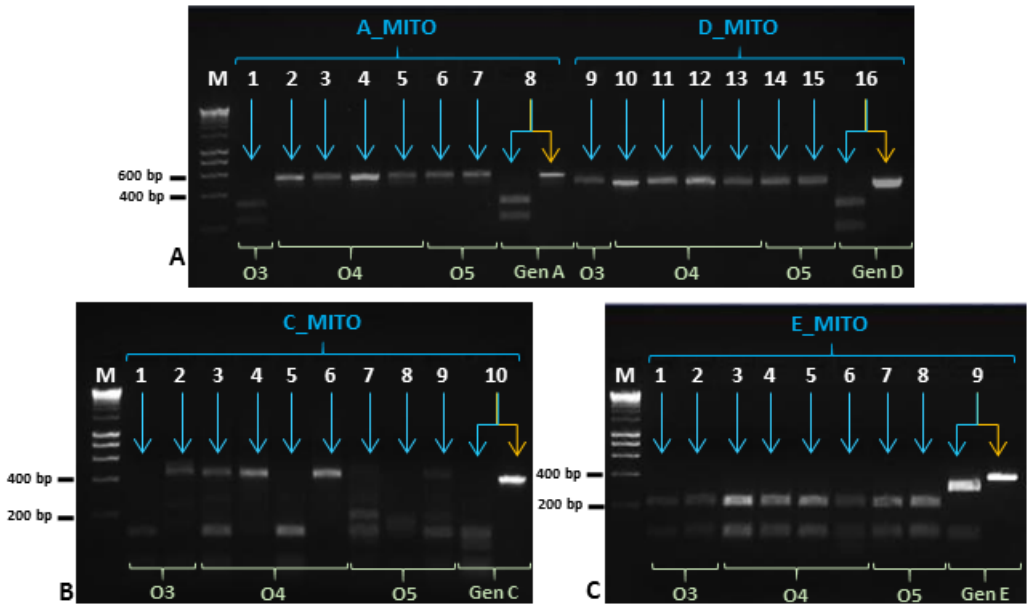


Figure 5.20 Outbreak samples O3 to O6 restriction digestion results. **A**: A_MITO and D_MITO primer pairs: 1, O3-2; 2, O4-1; 3, O4-2; 4, O4-6; 5, O4-7; 6, O5-1.1; 7, O5-3.1; 8, *A. astaci* Da (genotype A); 9, O3-1; 10, O4-1; 11, O4-2; 12, O4-6; 13, O4-7; 14, O5-1.1; 15, O5-3.1; 16, *A. astaci* Pc (genotype D). **B**: C_MITO primer pair: 1, O3-1; 2, O3-2; 3, O4-1; 4, O4-2; 5, O4-6; 6, O4-7; 7, O5-1.1; 8, O5-2.1; 9, O5-3.1; 10, *A. astaci* KV (genotype C). **C**: E_MITO primer pair: 1, O3-1; 2, O3-3; 3, O4-1; 4, O4-2; 5, O4-6; 6, O4-7; 7, O5-1.1; 8, O5-3.1; 9, *A. astaci* KB13 (genotype E). In all figures: M, Bioline HyperLadder I. Blue arrows: PCR product after enzymatic digestion. Yellow arrows: original PCR product. For outbreaks description see section 2.1.8.

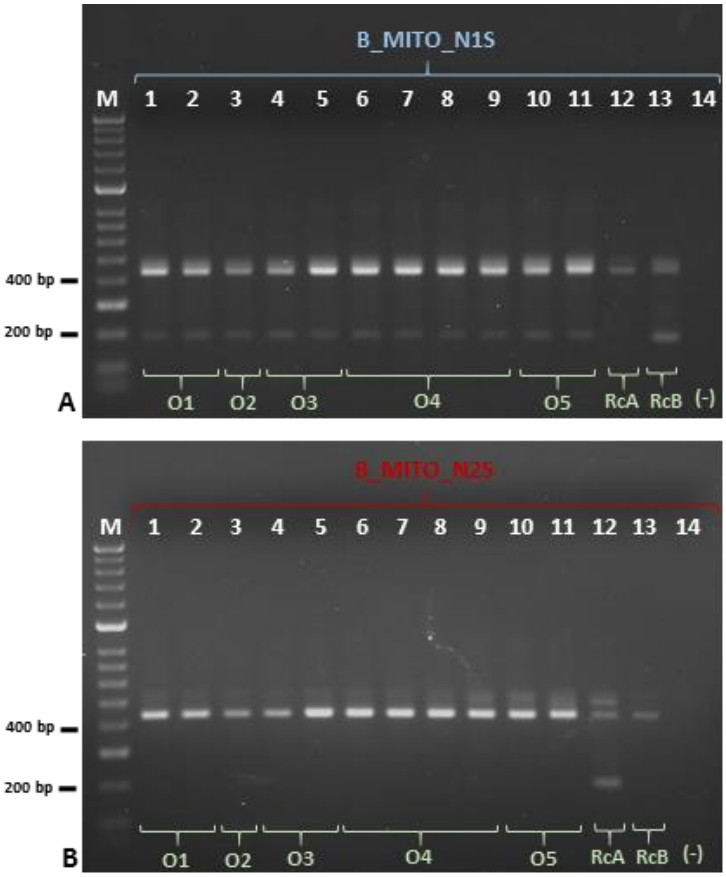


Figure 5.21 Outbreak samples O1 to O6 semi-nested PCR results. **A**, B_MITO_N1S; **B**, B_MITO_N2S. In both figures: M, Bioline HyperLadder II; 1, O1-1.2; 2, O1-1.2; 3, O2-1.2; 4, O3-1; 5, O3-3; 6, O4-1; 7, O4-2; 8, O4-6; 9, O4-7; 10, O5-1.1; 11, O5-3.1; 12, *A. astaci* Da (genotype A); 13, *A. astaci* D2 (genotype B); 14, negative control. RcA, PCR reaction control (genotype A); RcB, PCR reaction control (genotype B); (-), negative control. For outbreaks description see section 2.1.8.

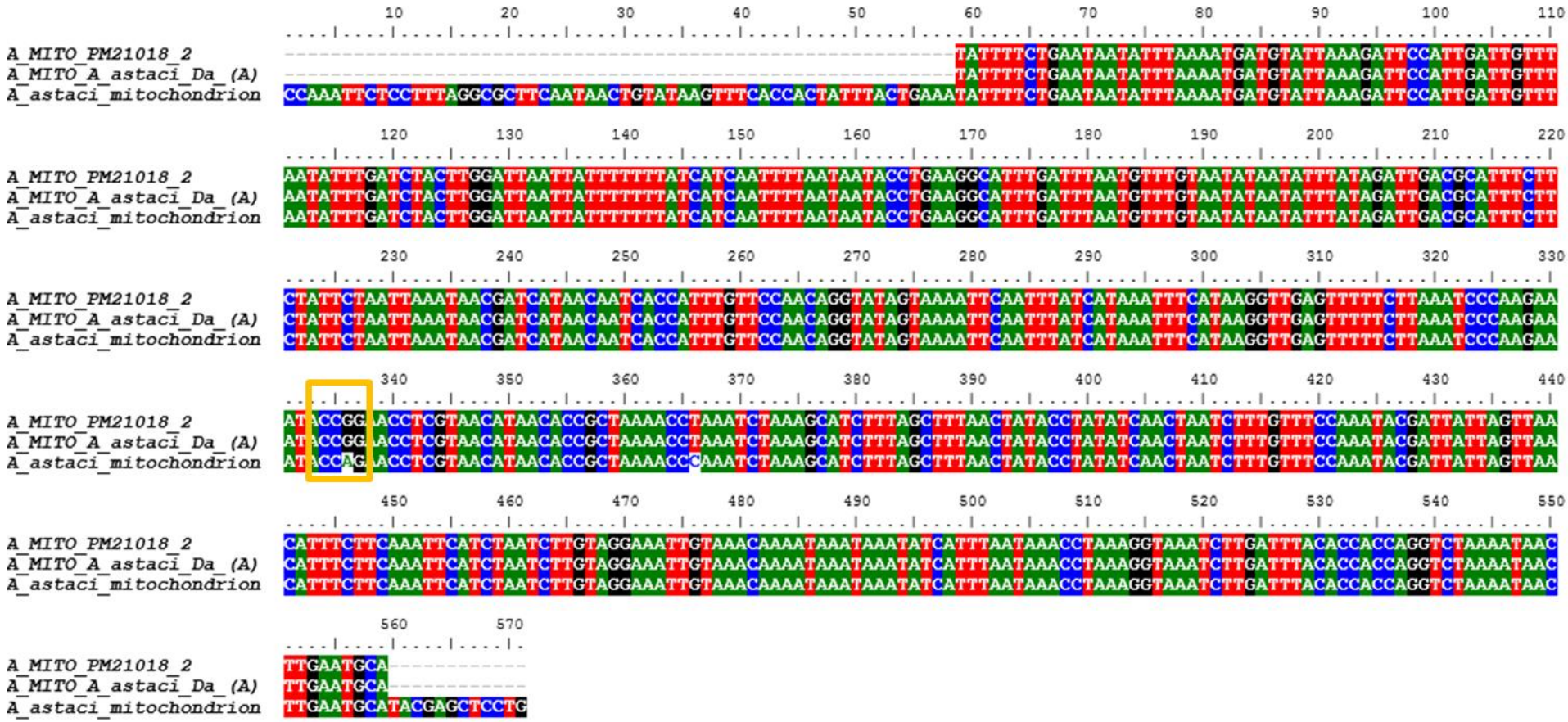


Figure 5.22 PM21018 sample number 2 PCR product amplified with A_MITO primer pair. The sequence and the SNV is matching the PCR product from *A. astaci* Da (genotype A). Yellow box: SNVs in restriciton sites. Last sequence: *A. astaci* mitochondrial reference sequence (genotype B).

5.10 Section 3.4.7

Table 5.13 PCR results from carriers samples amplified by A, B, C, D, and E mitochondrial primer pairs. +/orange: sample positive for PCR reaction; -/light blue: sample negative for PCR reaction.

		PCR 45 cycles with mitochondrial primer pairs on carriers samples									
Population	**Sample N**	**A_MITO**		**B_MITO**		**C_MITO**		**D_MITO**		**E_MITO**	
III	13-T	+	Figure 5.23 A well 1	+	Figure 5.24 A well 1	+	Figure 5.25 A well 1	-	Figure 5.26 A well 1	+	Figure 5.27 A well 1
III	28-C	-	Figure 5.23 A well 2	+	Figure 5.24 A well 2	+	Figure 5.25 A well 2	-	Figure 5.26 A well 2	N/A	
IV	30-T	-	Figure 5.23 A well 3	+	Figure 5.24 A well 3	+	Figure 5.25 A well 3	+	Figure 5.26 A well 3	+	Figure 5.27 A well 2
VII	26-T	-	Figure 5.23 A well 4	+	Figure 5.24 A well 4	+	Figure 5.25 A well 4	-	Figure 5.26 A well 4	-	Figure 5.27 A well 3
VIII	6-T	-	Figure 5.23 A well 16	+	Figure 5.24 A well 5	+	Figure 5.25 A well 5	-	Figure 5.26 A well 5	-	Figure 5.27 A well 4
IX	6-C	+	Figure 5.23 A well 5	+	Figure 5.24 A well 6	+	Figure 5.25 A well 6	+	Figure 5.26 A well 6	+	Figure 5.27 A well 5
XIV	16-C	-	Figure 5.23 A well 6	+	Figure 5.24 A well 7	+	Figure 5.25 A well 7	+	Figure 5.26 A well 7	+	Figure 5.27 A well 6
XIV	19-C	+	Figure 5.23 A well 7	+	Figure 5.24 A well 8	+	Figure 5.25 A well 8	+	Figure 5.26 A well 8	+	Figure 5.27 A well 7
XV	2-C	+	Figure 5.23 A well 8	+	Figure 5.24 A well 9	+	Figure 5.25 A well 9	+	Figure 5.26 A well 9	+	Figure 5.27 A well 8
XV	8-C	+	Figure 5.23 A well 9	+	Figure 5.24 A well 10	+	Figure 5.25 A well 10	+	Figure 5.26 A well 10	+	Figure 5.27 A well 9
XVI	2-C	-	Figure 5.23 A well 10	+	Figure 5.24 A well 11	-	Figure 5.25 A well 11	-	Figure 5.26 A well 11	-	Figure 5.27 A well 10
XVI	8-C	+	Figure 5.23 A well 11	+	Figure 5.24 A well 12	+	Figure 5.25 A well 12	+	Figure 5.26 A well 12	+	Figure 5.27 A well 11
XVII	6a1-C	-	Figure 5.23 A well 12	+	Figure 5.24 A well 13	+	Figure 5.25 A well 13	+	Figure 5.26 A well 13	+	Figure 5.27 A well 12
XVII	9a2-C	+	Figure 5.23 A well 13	+	Figure 5.24 A well 14	+	Figure 5.25 A well 14	+	Figure 5.26 A well 14	+	Figure 5.27 A well 13
XVIII	14a2-C	+	Figure 5.23 A well 14	+	Figure 5.24 A well 15	+	Figure 5.25 A well 15	+	Figure 5.26 A well 15	-	Figure 5.27 A well 14
XVIII	16a1-C	-	Figure 5.23 A	+	Figure 5.24 A	+	Figure 5.25 A	+	Figure 5.26 A	+	Figure 5.27 A


well 15
well 16
well 16
well 16
well 15

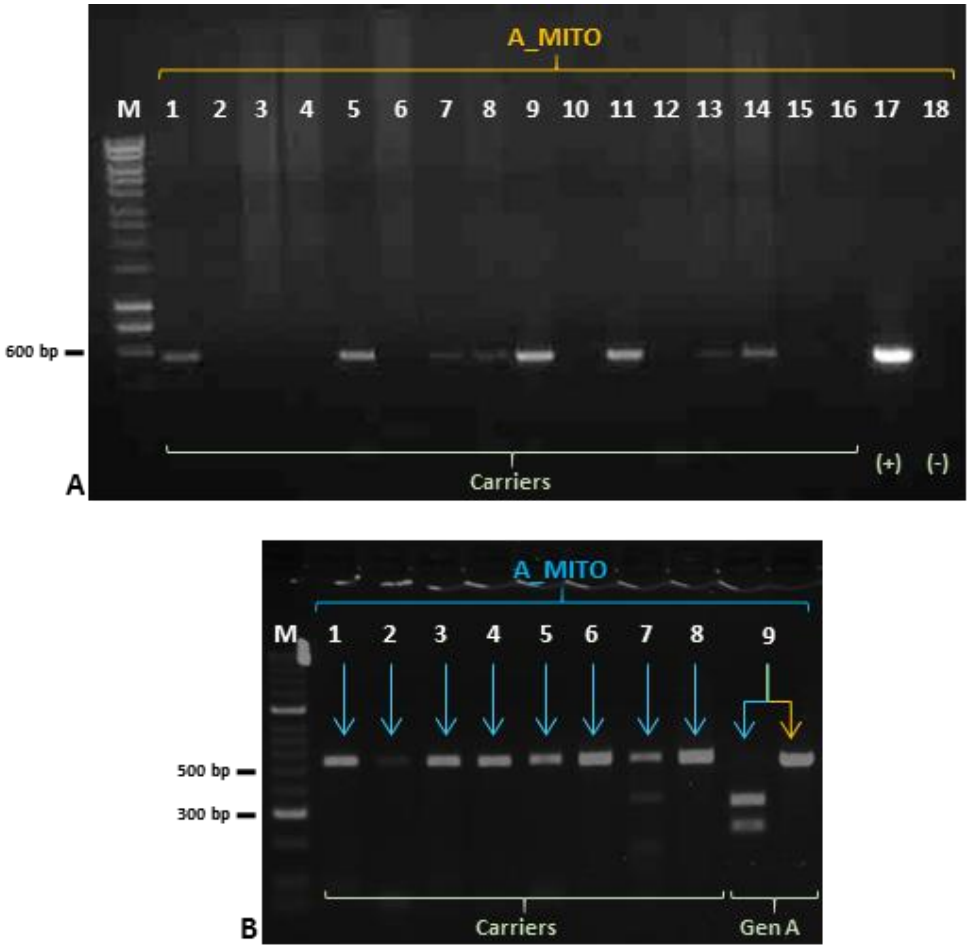


Figure 5.23 Carriers samples PCR (**A**) and enzymatic digestions (**B**) results for A_MITO primer pair. **A**: M, Bioline HyperLadder I; 1, III-13-T; 2, III-28-C; 3, IV-30-T; 4, VII-26-T; 5, IX-6-C; 6, XIV-16-C; 7, XIV-19-C; 8, XV-2-C; 9, XV-8-C; 10, XVI-2-C; 11, XIV-8-C; 12, XVII-6a1-C; 13, XVII-9a2-C; 14, XVIII-14a2-C; 15, XVIII-16a1-C; 16, VIII-6-T; 17, *A. astaci* Da (genotype A); 18, negative control. **B**: M, Bioline HyperLadder II; 1, III-13-T; 2, IX-6-C; 3, XIV-19-C; 4, XV-2-C; 5, XV-8-C; 6, XIV-8-C; 7, XVII-9a2-C; 8, XVIII-14a2-C; 9, *A. astaci* Da (genotype A). (+), positive control; (-), negative control. Blue arrows: PCR product after enzymatic digestion. Yellow arrows: original PCR product. For carriers description see section 2.1.8.

.

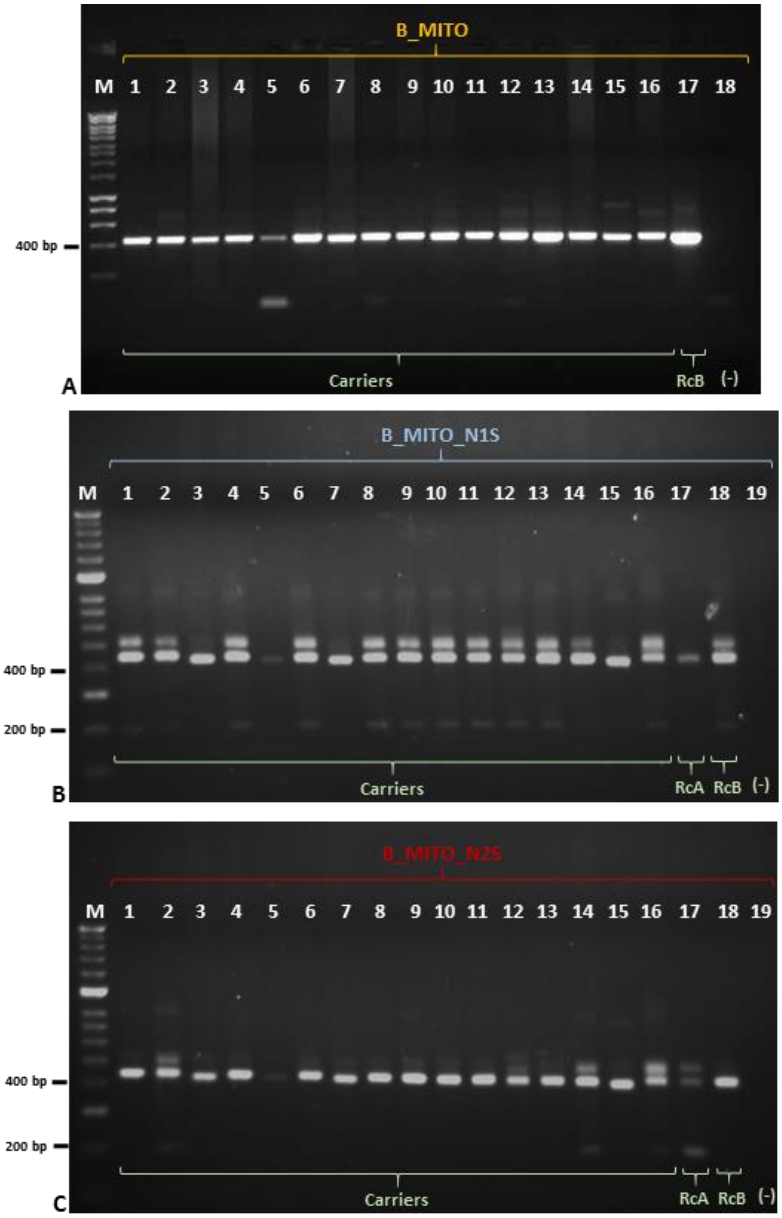


Figure 5.24 Carriers samples PCR with B_MITO primer pair (**A**), semi-nested PCR with B_MITO_N1S primer (**B**) and semi-nested PCR with B_MITO_N2S primer (**C**) results. **A**: M, Bioline HyperLadder I; 1, III-13-T; 2, III-28-C; 3, IV-30-T; 4, VII-26-T; 5, VIII-6-T; 6, IX-6-C; 7, XIV-16-C; 8, XIV-19-C; 9, XV-2-C; 10, XV-8-C; 11, XVI-2-C; 12, XIV-8-C; 13, XVII-6a1-C; 14, XVII-9a2-C; 15, XVIII-14a2-C; 16, XVIII-16a1-C; 17, *A. astaci* D2 (genotype B); 18, negative control. **B** and **C**: M, Bioline HyperLadder II; 1-16, semi-nested results on previous positive carriers; 17, *A. astaci* Da (genotype A); 18, *A. astaci* D2 (genotype B); 19, negative control. For carriers description see section 2.1.8.

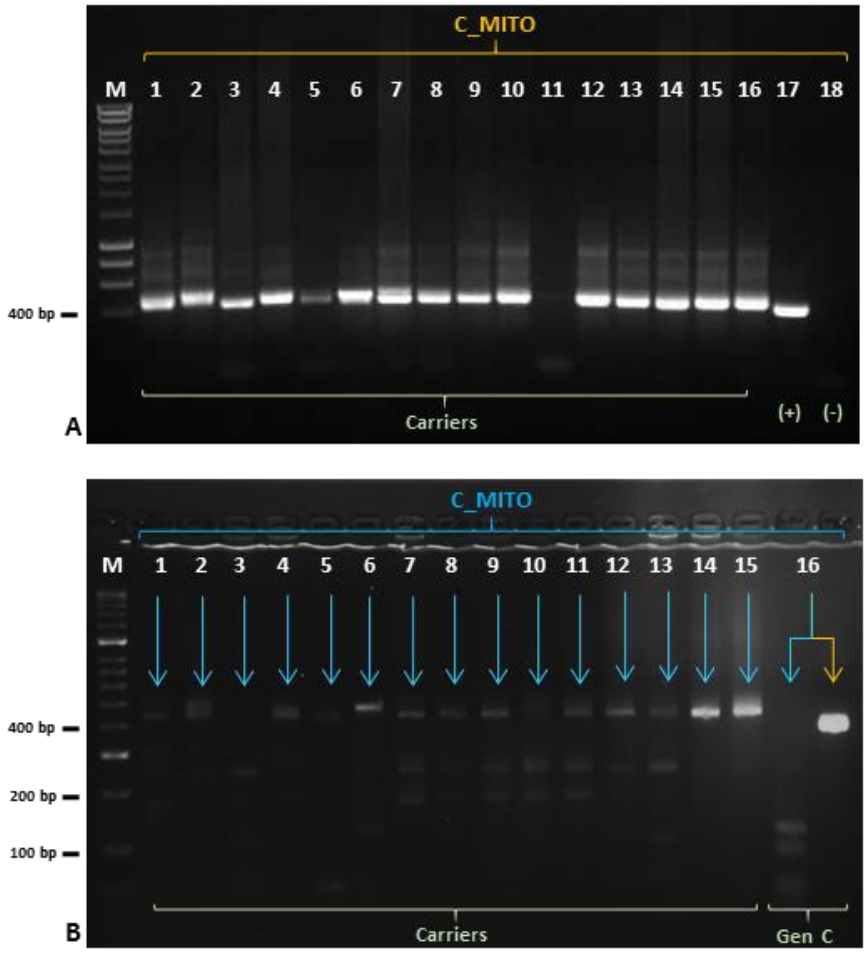


Figure 5.25 Carriers samples PCR (**A**) and restriction digestion (**B**) results for C_MITO primer pair. **A**: M, Bioline HyperLadder I; 1, III-13-T; 2, III-28-C; 3, IV-30-T; 4, VII-26-T; 5, VIII-6-T; 6, IX-6-C; 7, XIV-16-C; 8, XIV-19-C; 9, XV-2-C; 10, XV-8-C; 11, XVI-2-C; 12, XIV-8-C; 13, XVII-6a1-C; 14, XVII-9a2-C; 15, XVIII-14a2-C; 16, XVIII-16a1-C; 17, *A. astaci* KV (genotype C); 18, negative control. **B**: M, Bioline HyperLadder II; 1, III-13-T; 2, III-28-C; 3, IV-30-T; 4, VII-26-T; 5, VIII-6-T; 6, IX-6-C; 7, XIV-16-C; 8, XIV-19-C; 9, XV-2-C; 10, XV-8-C; 11, XIV-8-C; 12, XVII-6a1-C; 13, XVII-9a2-C; 14, XVIII-14a2-C; 15, XVIII-16a1-C; 16, *A. astaci* KV (genotype C). (+), positive control; (-), negative control. Blue arrows: PCR product after enzymatic digestion. Yellow arrows: original PCR product. For carriers description see section 2.1.8.

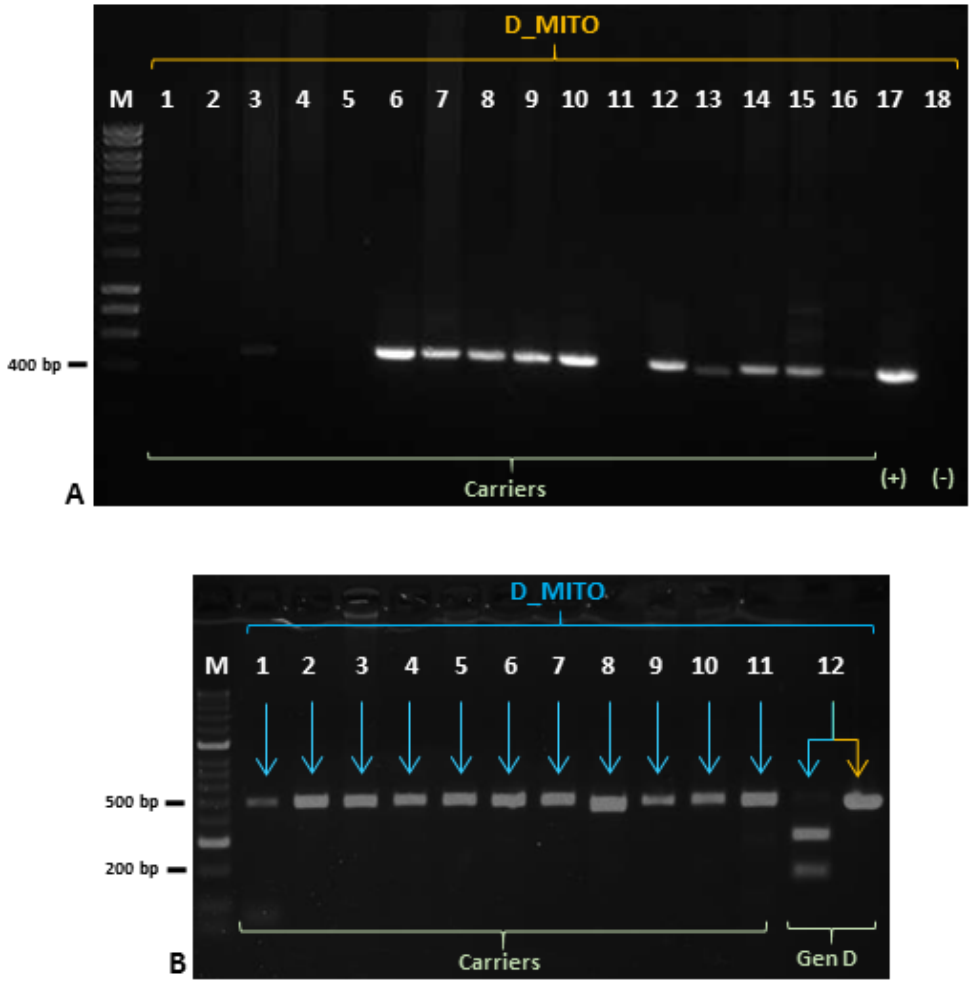


Figure 5.26 Carriers samples PCR (**A**) and restriction digestion (**B**) results for D_MITO primer pair. **A**: M, Bioline HyperLadder I; 1, III-13-T; 2, III-28-C; 3, IV-30-T; 4, VII-26-T; 5, VIII-6-T; 6, IX-6-C; 7, XIV-16-C; 8, XIV-19-C; 9, XV-2-C; 10, XV-8-C; 11, XVI-2-C; 12, XIV-8-C; 13, XVII-6a1-C; 14, XVII-9a2-C; 15, XVIII-14a2-C; 16, XVIII-16a1-C; 17, *A. astaci* Pc (genotype D); 18, negative control. **B**: M, Bioline HyperLadder II; 1, IV-30-T; 2, IX-6-C; 3, XIV-16-C; 4, XIV-19-C; 5, XV-2-C; 6, XV-8-C; 7, XIV-8-C; 8, XVII-6a1-C; 9, XVII-9a2-C; 10, XVIII-14a2-C; 11, XVIII-16a1-C; 12, *A. astaci* Pc (genotype D). (+), positive control; (-), negative control. Blue arrows: PCR product after enzymatic digestion. Yellow arrows: original PCR product. For carriers description see section 2.1.8.

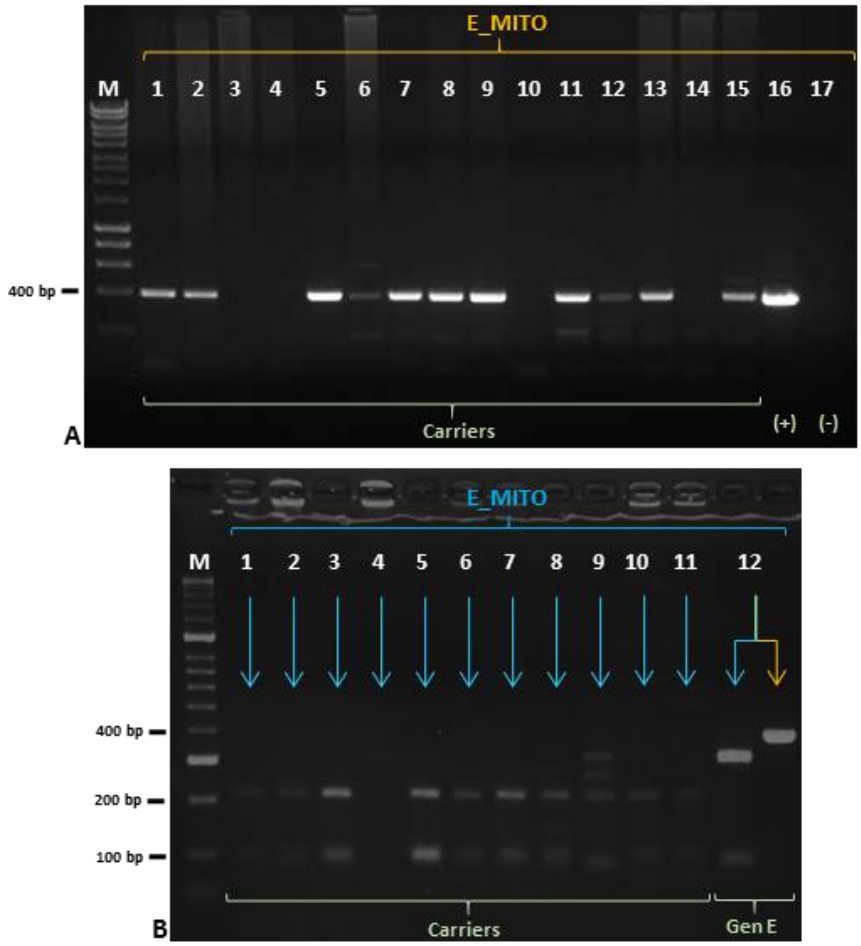


Figure 5.27 Carriers samples PCR (**A**) and restriction digestion (**B**) results for E_MITO primer pair. **A**: M, Bioline HyperLadder I; 1-15, III-13-T, IV-30-T, VII-26-T, VIII-6-T, IX-6-C, XIV-16-C, XIV-19-C, XV-2-C, XV-8-C, XVI-2-C, XIV-8-C, XVII-6a1-C, XVII-9a2-C, XVIII-14a2-C, XVIII-16a1-C; 16, *A. astaci* KB13; 17, negative control. **B**: M, Bioline HyperLadder II; 1-11, digested PCR products from previous positive carriers; 12, *A. astaci* KB13. (+), positive control; (-), negative control. Blue arrows: PCR product after enzymatic digestion. Yellow arrows: original PCR product. For carriers description see section 2.1.8.

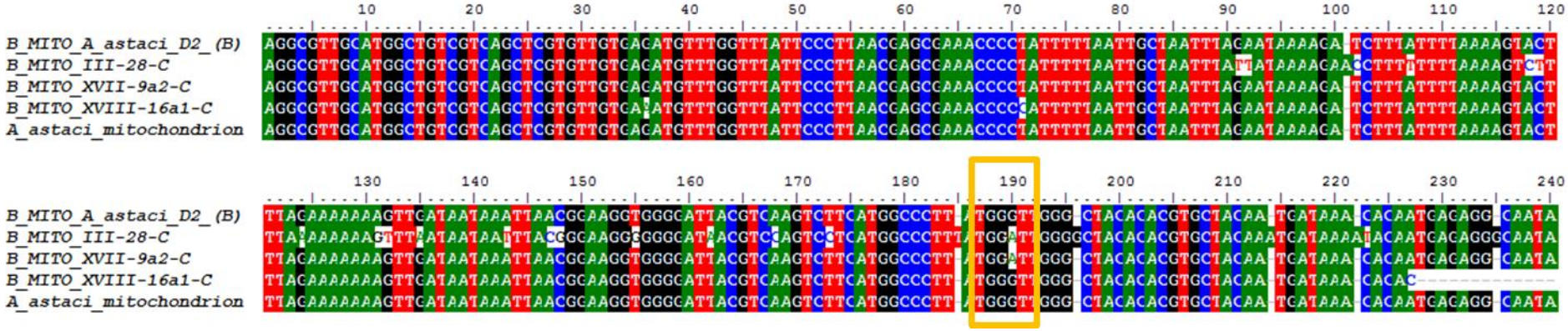


Figure 5.28 Sequences of PCR products amplified with the semi-nested PCR from 3 carriers samples. These three samples were both positive at the two B_MITO semi-nested primers. The SNV of samples III-28-C matched the ones from *A. astac*i D2 (genotype B) and *A. astaci* mitochondrial reference sequence (genotype B), the PCR product presents some differences in comparison to the reference sequence. The SNV of sample XVIII-16a1-C matched ones from *A. astaci* D2 (genotype B) and *A. astaci* mitochondrial reference sequence (genotype B), the PCR product present a highly similar sequence to the reference sequence. The SNV of sample XVII-9a2-C does not match the ones from *A. astaci* D2 (genotype B) and *A. astaci* mitochondrial reference sequence (genotype B), but the PCR product shared the same sequence. These alignments combined with the semi-nested PCR results indicate the presence of *A. astaci* genotype B and some other oomycete (or an unknown genotype) in the sample. Yellow box: SNV in Last sequence: *A. astaci* mitochondrial reference sequence (genotype B).

5.11 Section 3.5.3

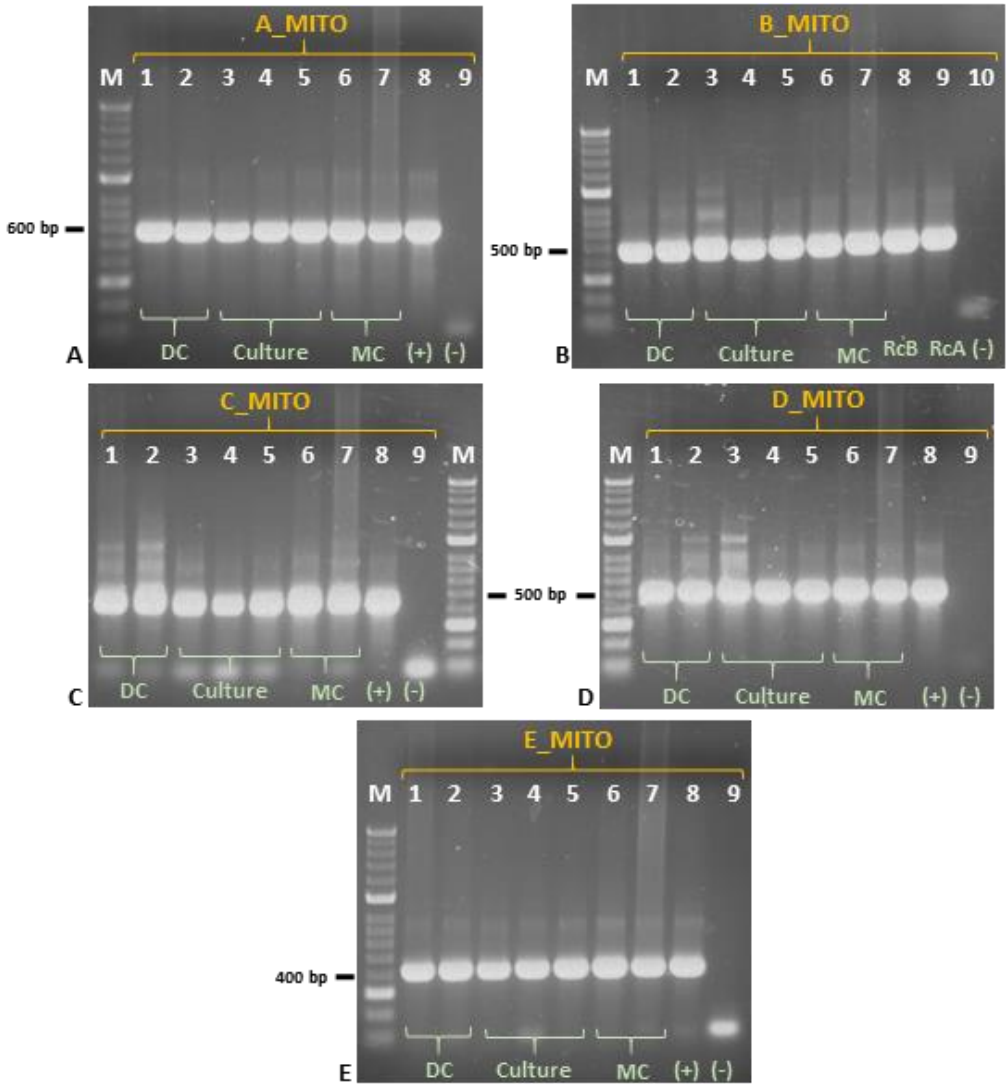


Figure 5.29 Genotyping an Italian crayfish plague outbreak with assays exploiting SNVs in the mtDNA. **A**, **B**, **C**, **D** and **E**: 1-2, gDNA extraction from dead *A. pallipes* cuticle; 3-5, gDNA extraction from *A. astaci* isolated during the outbreak; 6-7, gDNA extraction from moribund *A. pallipes* cuticle; **A**8, *A. astaci* Da (genotype A); **B**8-9, *A. astaci* D2 (genotype B) and Da (genotype A); **C**8, *A. astaci* KV (genotype C); **D**8, *A. astaci* Pc (genotype D); **E**8, *A. astaci* KB13 (genotype E); **A**9, **B**10, **C**9, **D**9 and **E**9, negative control; M, Bioline HyperLadder II. RcA, PCR reaction control (genotype A); RcB, PCR reaction control (genotype B); (+), positive control; (-), negative control.

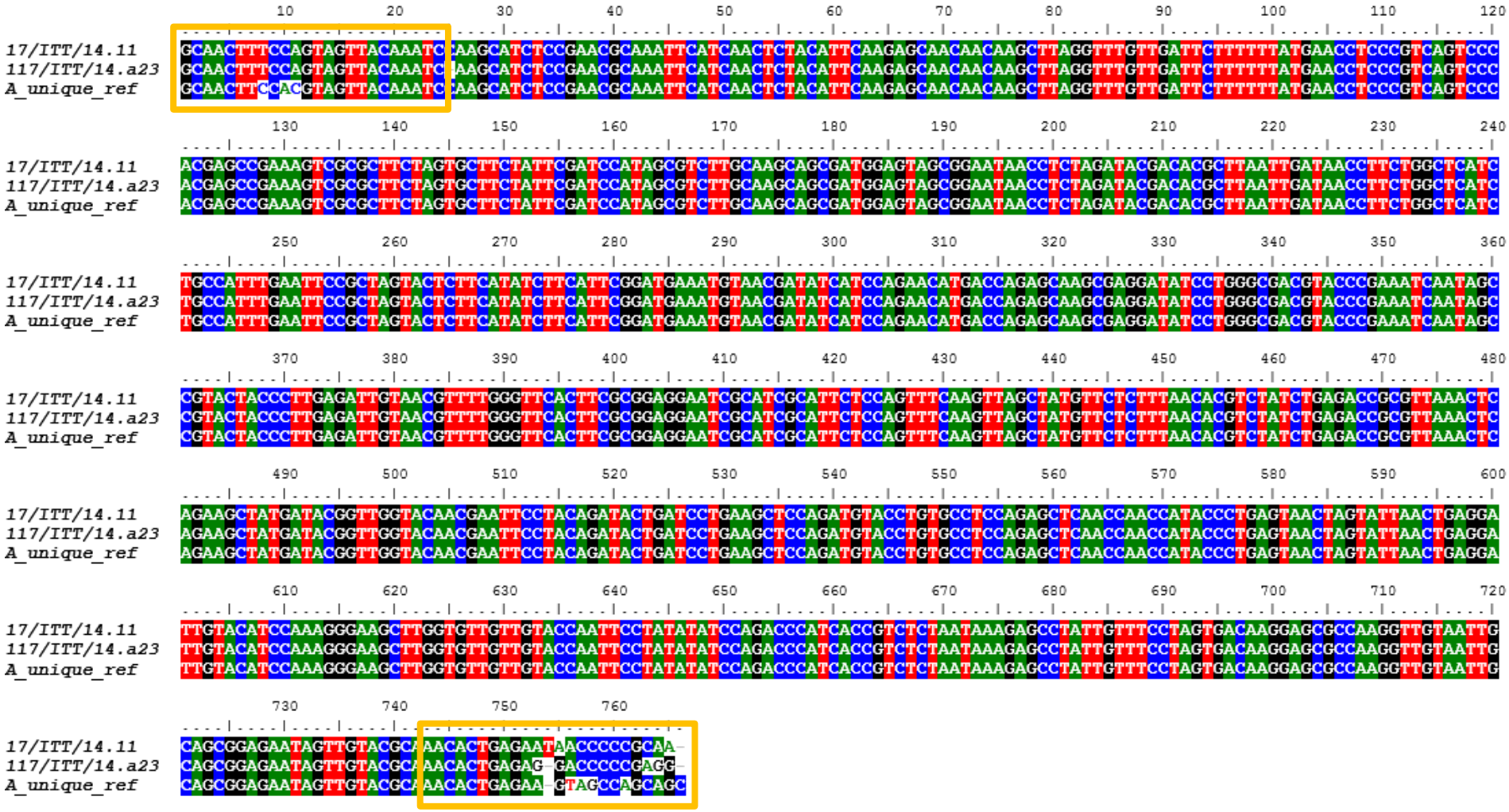


Figure 5.30 Alignment of sequenced PCR fragments amplified by primer pair A_unique from a subset of samples taken during the Italian outbreak. 17/ITT/14.11, cuticle of dead *A. pallipes*; 117/ITT/14.α23, *A. astaci* isolated during the outbreak; A_unique_ref, genotype A specific region reference sequence. Yellow boxes: primers sites.

5.12 Section 3.6.4

Table 5.14 Dnadiff percentage of sequence identity of all *A. astaci*, *A. invadans*, *A. invadans*-like, and *A. frigidophilus* isolates sequenced in this study, First column: *A. astaci* reference isolate; second column: isolates assemblies aligned against the reference; third column: percentage of sequence identity between the assemblies.

Isolate	vs isolate	% sequence identity		Isolate	vs isolate	% sequence identity
A. astaci 197901 (B)	A-Da	96.66		*A. astaci* Da (A)	A-SV	99.41
	A-SV	96.67			B-197901	96.66
	B-457	99.3			B-457	96.85
	B-D2	98.9			B-D2	96.4
	B-SA	98.96			B-SA	96.39
	B-Si	98.91			B-Si	96.35
	B-YX	98.92			B-YX	96.4
	C-KV	98.46			C-KV	96.31
	D-APO3	95.9			D-APO3	95.5
	D-Pc	95.89			D-Pc	95.52
	E-KB13	98.54			E-KB13	97.43
	9701*	85.04			9701*	84.98
	0002*	85.32			0002*	85.26
	9510**	89.1			9510**	88.74
	AP5***	85.34			AP5***	85.32
	RP2***	85.22			RP2***	85.25
A. astaci 457 (B)	A-Da	96.85		*A. astaci* SV (A)	A-Da	99.41
	A-SV	96.85			B-197901	96.67
	B-197901	99.3			B-457	96.85
	B-D2	98.95			B-D2	96.41
	B-SA	99.04			B-SA	96.39
	B-Si	98.9			B-Si	96.36
	B-YX	98.91			B-YX	96.4
	C-KV	98.43			C-KV	96.31
	D-APO3	96.04			D-APO3	95.52
	D-Pc	96.02			D-Pc	95.54
	E-KB13	98.62			E-KB13	97.43
	9701*	85.05			9701*	84.97
	0002*	85.33			0002*	85.26
	9510**	90.34			9510**	88.9
	AP5***	85.38			AP5***	85.32
	RP2***	85.24			RP2***	85.25
A. astaci D2 (B)	A-Da	96.4		*A. astaci* KV (C)	A-Da	96.31
	A-SV	96.41			A-SV	96.31
	B-197901	98.9			B-197901	98.46
	B-457	98.95			B-457	98.43
	B-SA	98.72			B-D2	98.18
	B-Si	98.75			B-SA	98.2
	B-YX	98.74			B-Si	98.22
	C-KV	98.18			B-YX	98.27
	D-APO3	95.66			D-APO3	95.68
	D-Pc	95.69			D-Pc	95.71
	E-KB13	98.15			E-KB13	98.18
	9701*	85.02			9701*	84.98
	0002*	85.31			0002*	85.36
	9510**	88.69			9510**	88.77
	AP5***	85.3			AP5***	85.32
	RP2***	85.23			RP2***	85.25
A. astaci SA (B)	A-Da	96.39		*A. astaci* APO3 (D)	A-Da	95.5
	A-SV	96.39			A-SV	95.52
	B-197901	98.96			B-197901	95.9
	B-457	99.04			B-457	96.04
	B-D2	98.72			B-D2	95.66
	B-Si	98.78			B-SA	95.69
	B-YX	98.72			B-Si	95.67
	C-KV	98.2			B-YX	95.66
	D-APO3	95.69			C-KV	95.68
	D-Pc	95.71			D-Pc	99.47
	E-KB13	98.17			E-KB13	96.14
	9701*	85.01			9701*	84.92
	0002*	85.3			0002*	85.11
	9510**	88.74			9510**	88.95
	AP5***	85.33			AP5***	85.21

	RP2***	85.22			RP2***	85.16
A. astaci Si (B)	A-Da	96.35		*A. astaci* Pc (D)	A-Da	95.52
	A-SV	96.36			A-SV	95.54
	B-197901	98.91			B-197901	95.89
	B-457	98.9			B-457	96.02
	B-D2	98.75			B-D2	95.69
	B-SA	98.78			B-SA	95.71
	B-YX	98.75			B-Si	95.69
	C-KV	98.22			B-YX	95.69
	D-APO3	95.67			C-KV	95.71
	D-Pc	95.69			D-APO3	99.47
	E-KB13	98.15			E-KB13	96.13
	9701*	85			9701*	85
	0002*	85.27			0002*	85.3
	9510**	88.74			9510**	88.8
	AP5***	85.33			AP5***	85.29
	RP2***	85.25			RP2***	85.23
A. astaci YX (B)	A-Da	96.4		*A. astaci* KB13 (E)	A-Da	97.43
	A-SV	96.4			A-SV	97.43
	B-197901	98.92			B-197901	98.54
	B-457	98.91			B-457	98.62
	B-D2	98.74			B-D2	98.15
	B-SA	98.72			B-SA	98.17
	B-Si	98.75			B-Si	98.15
	C-KV	98.27			B-YX	98.15
	D-APO3	95.66			C-KV	98.18
	D-Pc	95.69			D-APO3	96.14
	E-KB13	98.15			D-Pc	96.13
	9701*	85.03			9701*	85.32
	0002*	85.31			0002*	85.63
	9510**	88.7			9510**	89.38
	AP5***	85.31			AP5***	85.54
	RP2***	85.27			RP2***	85.44
A. invadans NJM9701	A-Da	84.98		*A. frigidophilus* AP5	A-Da	85.32
	B-D2	85.02			B-D2	85.3
	C-KV	84.98			C-KV	85.32
	D-Pc	85			D-Pc	85.29
	E-KB13	85.32			E-KB13	85.54
	0002*	99.16			9701*	85.24
	9510**	87.57			9510**	91.57
	AP5***	85.24			RP2***	96.13
A. invadans-like NJM9510	A-Da	88.74				
	B-D2	88.69				
	C-KV	88.77				
	D-Pc	88.8				
	E-KB13	89.38				
	9701*	87.57				
	AP5***	91.57				

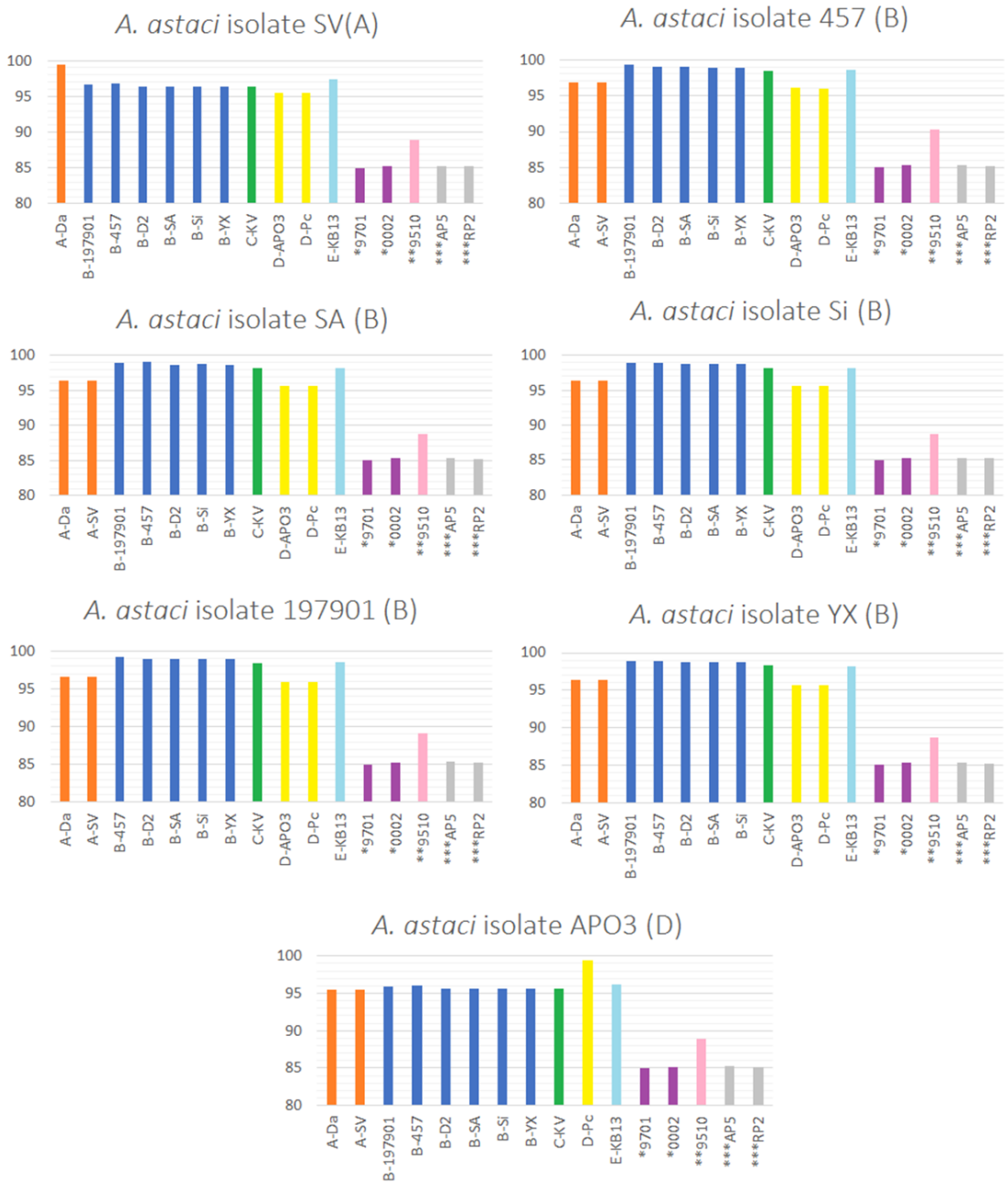


Figure 5.31 Remaining isolates dnadiff percentage of sequence identity. Y axis: % of sequence identity; X axis: isolates; orange bars: *A. astaci* SV and Da (genotype A); dark blue bars: *A. astaci* D2, Si, SA, YX, 457, and 197901 (genotype B); green bars: *A. astaci* KV (genotype C); yellow bars: *A. astaci* APO3 and Pc (genotype D); light blue bars: *A. astaci* KB13 (genotype E); purple bars: *A. invadans* NJM9701 and NJM0002; pink bars: *A. invadans*-like NJM9510; grey bars: *A. frigidophilus* AP5 and RP2. *, *A. invadans*. **, *A. invadans*-like. ***, *A. frigidophilus*.

6. Bibliography

Abdul-Muneer, P. M. (2014) Application of microsatellite markers in monservation genetics and fisheries management: recent advances in population structure analysis and conservation strategies. Genetics Research International, **2014**, 691759.

Acinas, S. G., Sarma-Rupavtarm, R., Klepac-Ceraj, V., *et al.* (2005) PCR-induced sequence artifacts and bias: insights from comparison of two 16S rRNA clone libraries constructed from the same sample. *Applied and Environmental Microbiology*, **71**, 8966–8969.

Ackefors, H. (1999) The positive effects of established crayfish introductions in Europe. In: *Crayfish in Europe as Alien Species —How to Make the Best of a Bad Situation?* (eds Gherardi, F., & Holdich, D. M.), pp. 49–61, Balkema, Rotterdam.

Ackermann, H.-W., & Kropinski, A. M. (2007) Curated list of prokaryote viruses with fully sequenced genomes. *Research in Microbiology*, **158**, 555–566.

Alderman, D. J. (1996) Geographical spread of bacterial and fungal diseases of crustaceans. *Revue Scientifique et Technique (International Office of Epizootics)*, **15**, 603–632.

Alderman, D. J. (2003) *Aphanomycosis of crayfish: crayfish plague. Research and Development Technical Report W2-064.* Cefas, Weymouth.

Alderman, D. J., Polglase, J. L., Frayling, M., *et al.* (1984) Crayfish plague in Britain. *Journal of Fish Diseases*, **7**, 401–405.

Alderman, D. J., Polglase, J. L., & Frayling, M. (1987) *Aphanomyces astaci* pathogenicity under laboratory and field conditions. *Journal of Fish Diseases*, **10**, 385–393.

Ali, B. A., Huang, T.-H., Qin, D.-N., *et al.* (2004) A review of random amplified polymorphic DNA (RAPD) markers in fish research. *Reviews in Fish Biology and Fisheries*, **14**, 443–453.

Álvarez, I., & Wendel, J. F. (2003) Ribosomal ITS sequences and plant phylogenetic inference. *Molecular Phylogenetics and Evolution*, **29**, 417–

434.

Anderson, I. J., Dharmarajan, L., Rodriguez, J., *et al.* (2009) The complete genome sequence of *Staphylothermus marinus* reveals differences in sulfur metabolism among heterotrophic Crenarchaeota. *BMC Genomics*, **10**, 145.

Andersson, S. G. E., Karlberg, O., Canbäck, B., *et al.* (2003) On the origin of mitochondria: a genomics perspective. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **358**, 165–179.

Andrews, S. (2010) FastQC: a quality control tool for high throughput sequence data. *Available Online at: Http://www.bioinformatics.babraham.ac.uk/projects/fastqc*.

Aydin, H., Kokko, H., Makkonen, J., *et al.* (2014) The signal crayfish is vulnerable to both the As and the PsI-isolates of the crayfish plague. *Knowledge and Management of Aquatic Ecosystems*, **413**, 3.

Baker, M. (2012) De novo genome assembly: what every biologist should know. *Nature Methods*, **9**, 333–337.

Bakkeren, G., Kronstad, J. W., & Lévesque, C. A. (2000) Comparison of AFLP fingerprints and ITS sequences as phylogenetic markers in Ustilaginomycetes. *Mycologia*, **92**, 510–521.

Bala, K., Robideau, G. P., Désaulniers, N., *et al.* (2010) Taxonomy, DNA barcoding and phylogeny of three new species of *Pythium* from Canada. *Persoonia*, **25**, 22–31.

Ballesteros, I., Martín, M. P., Cerenius, L., *et al*. (2007) Lack of specificity of the molecular diagnostic method for identification of *Aphanomyces. Bulletin Français de la Pêche et de la Pisciculture,* **385**, 17–24.

Bankevich, A., Nurk, S., Antipov, D., *et al.* (2012) SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *Journal of Computational Biology*, **19**, 455–477.

Baum, D. A., & Shaw, K. L. (1995) Genealogical perspectives on the species problem. In: *Experimental and molecular approaches to plant biosystematics* (eds Hoch, P. C., & Stephenson, A. G.), pp 289–303.

Missouri Botanical Garden, St. Louis.

Barraclough, T. G., Birky, C. W., & Burt, A. (2003) Diversification in sexual and asexual organisms. *Evolution*, **57**, 2166–2172.

Becking, T., Mrugała, A., Delaunay, C., *et al.* (2015) Effect of experimental exposure to differently virulent *Aphanomyces astaci* strains on the immune response of the noble crayfish *Astacus astacus*. *Journal of Invertebrate Pathology*, **132**, 115–124.

Birch, P. R. J., Armstrong, M., Bos, J., *et al.* (2009) Towards understanding the virulence functions of RXLR effectors of the oomycete plant pathogen *Phytophthora infestans*. *Journal of Experimental Botany*, **60**, 1133–1140.

Birky, C. W., Adams, J., Gemmel, M., *et al.* (2010) Using population genetic theory and DNA sequences for species detection and identification in asexual organisms. *PLoS ONE,* **5**, e10609.

Birky C. W., Wolf, C., Maughan, H., *et al.* (2005) Speciationand selection without sex. *Hydrobiologia*, **546**, 29–45.

Birol, I., Raymond, A., Jackman, S. D., *et al.* (2013) Assembling the 20 Gb white spruce (*Picea glauca*) genome from whole-genome shotgun sequencing data. *Bioinformatics*, **29**, 1492–1497.

Bos, J. (2007) Oomycete RXLR effectors: structural basis of avirulence and virulence in *Phytophthora infestans* AVR3a. *Phytopathology*, **97**, S145–S145.

Boutet, G., Alves Carvalho, S., Falque, M., *et al.* (2016) SNP discovery and genetic mapping using genotyping by sequencing of whole genome genomic DNA from a pea RIL population. *BMC Genomics*, **17**, 121.

Bruns, T. D., White, T. J., & Taylor, J. W. (1991) Fungal molecular systematics. *Annual Review of Ecology and Systematics*, **22**, 525–564.

Buller, N. B. (2014) *Bacteria and fungi from fish and other aquatic animals : a practical identification manual*, 2nd editio. Wallingford, Oxfordshire ; Boston, MA : CABI, [2014].

Caprioli, R., Cargini, D., Marcacci, M., *et al.* (2013) Self-limiting outbreak of crayfish plague in an *Austropotamobius pallipes* population of a river basin in the Abruzzi region (central Italy). *Diseases of Aquatic Organisms*, **103**, 149–156.

Carneiro Vieira, M. L., Santini, L., Lima Diaz, A., *et al*. (2016) Microsatellite markers: what they mean and why they are so useful. *Genetics and Molecular Biology*, **39**, 312–328.

Cerenius, L., Bangyeekhun, E., Keyser, P., *et al.* (2003) Host prophenoloxidase expression in freshwater crayfish is linked to increased resistance to the crayfish plague fungus, *Aphanomyces astaci*. *Cellular Microbiology*, **5**, 353–357.

Cerenius, L., Lee, B. L., & Söderhäll, K. (2008) The proPO-system: pros and cons for its role in invertebrate immunity. *Trends in Immunology*, **29**, 263–271.

Cerenius, L., & Söderhäll, K. (1984) Chemotaxis in *Aphanomyces aslaci,* an arthropod parasite. *Journal of Invertebrate Pathology*, **43**, 278–281.

Cerenius, L., & Söderhäll, K. (1985) Repeated Zoospore Emergence to Parasitism as a Possible Adaptation in *Aphanomyces*. *Experimental Mycology*, **9**, 259–263.

Chen, X., & Sullivan, P. F. (2003) Single nucleotide polymorphism genotyping: biochemistry, protocol, cost and throughput. *Pharmacogenomics Journal*, **3**, 77–96.

Clarke, C. R., Studholme, D. J., Hayes, B., *et al.* (2015) Genome-Enabled Phylogeographic Investigation of the Quarantine Pathogen *Ralstonia solanacearum* Race 3 Biovar 2 and Screening for Sources of Resistance Against Its Core Effectors. *Phytopathology*, **105**, 597–607.

Cline, J., Braman, J. C., & Hogrefe, H. H. (1996) PCR fidelity of pfu DNA polymerase and other thermostable DNA polymerases. *Nucleic Acids Research*, **24**, 3546–3551.

Cooke, D. E., Drenth, A., Duncan, J. M., *et al.* (2000) A molecular phylogeny of

Phytophthora and related oomycetes. *Fungal Genetics and Biology*, **30**, 17–32.

Cornalia, E. (1860) Sulla malattia dei gamberi. *Atti Della Società Italiana Di Scienze Naturali*, **2**, 334–336.

Cornish-Bowden, A. (1985) Nomenclature for incompletely specified bases in nucleic acid sequences: recommendations 1984. *Nucleic Acids Research*, **13**, 3021–3030.

Dakin, E. E., & Avise, J. C. (2004) Microsatellite null alleles in parentage analysis. *Heredity*, **93**, 504–509.

de Magalhães, J. P., Finch, C. E., & Janssens, G. (2010) Next-generation sequencing in aging research: emerging applications, problems, pitfalls and possible solutions. *Ageing Research Reviews*, **9**, 315–323.

Denoeud, F., Roussel, M., Noel, B., *et al.* (2011) Genome sequence of the stramenopile *Blastocystis*, a human anaerobic parasite. *Genome Biology*, **12**, R29.

Diéguez-Uribeondo, J., Cerenius, L., Dyková, I., *et al.* (2006) Pathogens, parasites and ectocommensals. In: *Atlas of Crayfish in Europe* (eds Souty-Grosset, C., Holdich, D. M., Noël, P. Y., et al.), Vol. 64, pp. 133–149. Colleciton Patrimoines Naturels, Paris: Muséum National d'Histoire Naturelle.

Diéguez-Uribeondo, J., García, M. A., Cerenius, L., *et al.* (2009) Phylogenetic relationships among plant and animal parasites, and saprotrophs in *Aphanomyces* (Oomycetes). *Fungal Genetics and Biology*, **46**, 365–376.

Diéguez-Uribeondo, J., Huang, T., Cerenius, L., *et al.* (1995) Physiological adaptation of an *Aphanomyces astaci* strain isolated from the freshwater crayfish *Procambarus clarkii*. *Mycological Research*, **99**, 574–578.

Diéguez-Uribeondo, J., Temino, C., & Muzquiz, J. L. (1997) The crayfish plague fungus (*Aphanomyces astaci*) in Spain. *Bulletin Français de La Pêche et de La Pisciculture*, **347**, 753–763.

Dobrowolski, M. P., Tommerup, I. C., Blakeman, H. D., *et al.* (2002) Non-

Mendelian inheritance revealed in a genetic analysis of sexual progeny of *Phytophthora cinnamomi* with microsatellite markers. *Fungal Genetics and Biology*, **35**, 197–212.

Duncan, J. M. (1999) *Phytophthora*—an abiding threat to our crops. *Microbiology Today*, **26**, 114–116.

Dykstra, M. J., Noga, E. J., Levine, J. F., *et al.* (1986) Characterization of the *Aphanomyces* species involved with ulcerative mycosis (UM) in menhaden. *Mycologia*, **78**, 664–672.

Edgar, R. C. (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, **32**.

Edgerton, B. F., Henttonen, P., Jussila, J., *et al.* (2004) Understanding the Causes of Disease in European Freshwater Crayfish. *Conservation Biology*, **18**, 1466–1474.

Edgerton, B., Paasonen, P., Henttonen, P., *et al.* (1996) Description of a bacilliform virus from the freshwater crayfish, *Astacus astacus*. *Journal of Invertebrate Pathology*, **68**, 187–190.

Edsman, L. (2004). The Swedish story about import of live crayfish. *Bulletin Français de la Pêche et de la Pisciculture*, **372**–**373**, 281–288.

Ernster, L., & Schatz, G. (1981) Mitochondria: a historical review. *The Journal of Cell Biology*, **91**, 227–255.

Fell, J. W., Boekhout, T., Fonseca, A., *et al.* (2000) Biodiversity and systematics of basidiomycetous yeasts as determined by large-subunit rDNA D1/D2 domain sequence analysis. *International Journal of Systematic and Evolutionary Microbiology,* **50**, 1351–1371.

Filipová, L., Petrusek, A., Matasová, K., *et al.* (2013) Prevalence of the Crayfish Plague Pathogen *Aphanomyces astaci* in Populations of the Signal Crayfish *Pacifastacus leniusculus* in France: Evaluating the Threat to Native Crayfish. *PLoS One*, **8**, e70157.

Fitch, W. M. (1970) Distinguishing homologous from analogous proteins. *Systematic Zoology*, **19**, 99–113.

Frazer, K. A., Elnitski, L., Church, D. M., *et al*. (2003) Cross-Species Sequence Comparisons: A Review of Methods and Available Resources. Genome Research, **13**, 1–12.

Fry, W. E., & Goodwin S. B. (1997) Resurgence of the Irish potato famine fungus. *Bioscience*, **47**, 363–371.

Gaulin, E., Jacquet, C., Bottin, A., *et al.* (2007) Root rot disease of legumes caused by *Aphanomyces euteiches*. *Molecular Plant Pathology*, **8**, 539–548.

Gaulin, E., Madoui, M.-A., Bottin, A., *et al.* (2008) Transcriptome of *Aphanomyces euteiches*: new oomycete putative pathogenicity factors and metabolic pathways. *PLoS ONE,* **3**, e1723.

Ganley, A. R., & Kobayashi, T. (2007) Highly efficient concerted evolution in the ribosomal DNA repeats: total rDNA repeat variation revealed by whole-genome shotgun sequence data. *Genome Research*, **17**, 184–191.

Gemayel, R., Cho, J., Boeynaems, S., *et al*. (2012) Beyond Junk-Variable Tandem repeats as facilitators of rapid evolution of regulatory and coding sequences. *Genes,* **3**, 461–480.

Grad, Y. H., Lipsitch, M., Feldgarden, M., *et al.* (2012) Genomic epidemiology of the *Escherichia coli* O104:H4 outbreaks in Europe, 2011. *Proceedings of the National Academy of Sciences*, **109**, 3065–3070.

Grandjean, F., Vrålstad, T., Diéguez-Uribeondo, J., *et al.* (2014) Microsatellite markers for direct genotyping of the crayfish plague pathogen *Aphanomyces astaci* (Oomycetes) from infected host tissues. *Veterinary Microbiology*, **170**, 317–324.

Gras, E., Matias-Guiu, X., Catasus, L., *et al.* (2000) Application of microsatellite PCR techniques in the identification of mixed up tissue specimens in surgical pathology. *Journal of Clinical Pathology,* **53**, 238–240.

Grenville-Briggs, L., Gachon, C. M., Strittmatter, M., *et al. (*2011) A molecular insight into algal-oomycete warfare: cDNA analysis of *Ectocarpus siliculosus* infected with the basal oomycete *Eurychasma dicksonii*. *PLoS*

ONE, **6**, e24500.

Gudbjartsson, D. F., Helgason, H., Gudjonsson, S. A., *et al.* (2015) Large-scale whole-genome sequencing of the Icelandic population. *Nature Genetics*, **47**, 435–444.

Gurevich, A., Saveliev, V., Vyahhi, N., *et al.* (2013) QUAST: quality assessment tool for genome assemblies. *Bioinformatics*, **29**, 1072–1075.

Haas, B. J., Kamoun, S., Zody, M.C., *et al*. (2009). Genome sequence and analysis of the Irish potato famine pathogen *Phytophthora infestans*. *Nature* **461**, 393–398.

Hall, T. A. (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series*, **41**, 95–98.

Halldén, C., Hansen, M., Nilsson, N.-O., *et al.* (1996) Competition as a source of errors in RAPD analysis. *Theoretical and Applied Genetics*, **93**, 1185–1192.

Hao, W., Allen, V. G., Jamieson, F. B., *et al.* (2012) Phylogenetic incongruence in *E. coli* O104: understanding the evolutionary relationships of emerging pathogens in the face of homologous recombination. *PLoS One*, **7**, e33971.

Harlioğlu, M. M. (2004) The present situation of freshwater crayfish, *Astacus leptodactylus* (Eschscholtz, 1823) in Turkey. *Aquaculture*, **230**, 181–187.

Harlioğlu, M. M., & Harlioğlu A. G. (2006) Threat of non-native crayfish introductions into Turkey: global lessons. *Reviews in Fish Biology and Fisheries,* **16**, 171–181.

Harris, S. A. (1999) RAPDs in systematics - a useful methodology? In: *Molecular Systematics and Plant Evolution* (eds Hollingsworth, P. M., Bateman, R. M., & Gornall, R. J.), p. 485. London: Taylor & Francis.

Harrison, R. G., & Larson, E. L. (2014) Hybridization, introgression, and the nature of species boundaries. *Journal of Heredity*, **105**, 795–809.

Hebert, P. D., Cywinska, A., Ball, S. L., *et al.* (2003) Biological identifications

through DNA barcodes. *Proceedings of the Royal Society of London. Series B, Biological Sciences*, **270**, 313–321.

Hehenberger, M. (2015) Nanomedicine: Science, Business, and Impact. Pan Stanford.

Hochwimmer, G., Tober, R., Bibars-Reiter, R., *et al.* (2009) Identification of two GH18 chitinase family genes and their use as targets for detection of the crayfish-plague oomycete *Aphanomyces astaci*. *BMC Microbiol*, **9**, 184.

Holdich, D. M. (1993) A review of *astaci*culture: freshwater crayfish farming *Aquatic Living Resources*, 6, 307–317.

Holdich, D.M., Reynolds, J.D., Souty-Grosset, C., *et al*. (2009) A review of the ever increasing threat to European crayfish from non-indigenous crayfish species. K*nowledge and Management of Aquatic Ecosystems,* **394/395**: 11.

Hsieh, C.-Y., Huang, C.-W., & Pan, Y.-C. (2016) Crayfish plague *Aphanomyces astaci* detected in redclaw crayfish, *Cherax quadricarinatus* in Taiwan. *Journal of Invertebrate Pathology*, **136**, 117–123.

Hu, J., Diao, Y., Zhou, Y., *et al.* (2013) Loss of heterozygosity drives clonal diversity of *Phytophthora capsici* in China. *PLoS One*, **8**, e82691.

Huang, T., Cerenius, L., & Söderhäll, K. (1994) Analysis of genetic diversity in the crayfish plague fungus, *Aphanomyces astaci*, by random amplification of polymorphic DNA. *Aquaculture*, **126**, 1–9.

Huson, D. H., & Bryant, D. (2006) Application of Phylogenetic Networks in Evolutionary Studies. *Molecular Biology and Evolution*, **23**, 254–267.

IHGSC, (International Human Genome Sequencing Consortium) (2004) Finishing the euchromatic sequence of the human genome. *Nature*, **431**, 931–945.

Ivanova, N., Sorokin, A., Anderson, I., *et al.* (2003) Genome sequence of *Bacillus cereus* and comparative analysis with *Bacillus anthracis*. *Nature*, **423**, 87–91.

James, J., Nutbeam-Tuffs, S., Cable, J., *et al.* (2017) The prevalence of *Aphanomyces astaci* in invasive signal crayfish from the UK and implications for native crayfish conservation. *Parasitology*, **144**, 411–418.

Jarne, P., & Lagoda, P. J. L. (1996) Microsatellites, from molecules to populations and back. *Trends in Ecology & Evolution*, **11**, 424–429.

Jiang, R. H., de Bruijn, I., Haas, B. J., *et al.* (2013) Distinctive expansion of potential virulence genes in the genome of the oomycete fish pathogen *Saprolegnia parasitica*. *PLoSGenetics*, **9**, e1003272.

Jiang, R. H., Tripathy, S., Govers, F., *et al.* (2008). RXLR effector reservoir in two *Phytophthora* species is dominated by a single rapidly evolving superfamily with more than 700 members. *Proceedings of the National Academy of Sciences of the United States of America*, **105**, 4874–4879.

Jiang, R. H. Y., Tyler, B. M., Whisson, S. C., *et al*. (2006). Ancient origin of elicitin gene clusters in *Phytophthora* genomes. *Molecular Biology Evolution,* **23**, 338–351.

Johannsen, W. (1911) The genotype conception of heredity. *The American Naturalist*, **45**, 129–159.

Jones, C. J., Edwards, K. J., Castaglione, S., *et al.* (1997) Reproducibility testing of RAPD, AFLP and SSR markers in plants by a network of European laboratories. *Molecular Breeding*, **3**, 381–390.

Jussila, J., Kokko, H., Kortet, R., *et al.* (2013) *Aphanomyces astaci* PsI-genotype isolates from different Finnish signal crayfish stocks show variation in their virulence but still kill fast. *Knowledge and Management of Aquatic Ecosystems*, 10.

Jussila, J., Maguire, I., Kokko, H., *et al*. (2015) Chaos and Adaptation in the Pathogen-Host Relationship in Relation to the Conservation: The Case of the Crayfish Plague and the Noble Crayfish. In: *Freshwater Crayfish, A Global Overview* (eds Kawai, T., Faulkes, Z., & Scholtz, G.), pp. 246–274. CRC Press.

Jussila, J., Makkonen, J., Vainikka, A., *et al.* (2011) Latent crayfish plague

(*Aphanomyces astaci*) infection in a robust wild noble crayfish (*Astacus astacus*) population. *Aquaculture*, **321**, 17–20.

Jussila, J., & Mannonen, A. (2004) Crayfisheries in Finland, a short overview. *Bulletin Français de La Pêche et de La Pisciculture*, **372–373**, 263–273.

Jussila, J., Vrezec, A., Makkonen, J., *et al.* (2016) Invasive Crayfish and Their Invasive Diseases in Europe with the Focus on the Virulence Evolution of the Crayfish Plague. In: *Biological Invasions in Changing Ecosystems, Vectors, Ecological Impacts, Management and Predictions* (ed Canning-Clode, J.), pp. 183–211. Berlin: De Gruyter Open.

Kajitani, R., Toshimoto, K., Noguchi, H., *et al.* (2014) Efficient de novo assembly of highly heterozygous genomes from whole-genome shotgun short reads. *Genome Research*, **24**, 1384–1395.

Kamoun, S. (2003) Molecular genetics of pathogenic oomycetes. *Eukaryot Cell*, **2**, 191–199.

Kamoun, S., Furzer, O., Jones, J. D. G., *et al*. (2015) The Top 10 oomycete pathogens in molecular plant pathology. *Molecular Plant Pathology*, **16**, 413–434.

Kasuga, T., Bui, M., Bernhardt, E., *et al.* (2016) Host-induced aneuploidy and phenotypic diversification in the Sudden Oak Death pathogen *Phytophthora ramorum*. *BMC Genomics*, **17**, 385.

Kim, C. S., Kosuke, Z., Nam, Y. K., *et al.* (2007) Protection of shrimp (*Penaeus chinensis*) against white spot syndrome virus (WSSV) challenge by double-stranded RNA. *Fish & Shellfish Immunology*, **23**, 242–246.

Kim, S.-R., Ramos, J., Ashikari, M., *et al.* (2016) Development and validation of allele-specific SNP/indel markers for eight yield-enhancing genes using whole-genome sequencing strategy to increase yield potential of rice, *Oryza sativa* L. *Rice*, **9**, 12.

Kitancharoen, N., & Hatai, K. (1997) *Aphanomyces frigidophilus* sp. nov. from eggs of Japanese char, *Salvelinus leucomaenis*. *Mycoscience*, **38**, 135–140.

Kong, Z., Zhao, P., Liu, H., *et al.* (2016) Whole-Genome Sequencing for the Investigation of a Hospital Outbreak of MRSA in China. *PLOS ONE*, **11**, e0149844.

Köser, C. U., Holden, M. T. G., Ellington, M. J., *et al.* (2012) Rapid Whole-Genome Sequencing for Investigation of a Neonatal MRSA Outbreak. *New England Journal of Medicine*, **366**, 2267–2275.

Kozubíková, E., Filipová, L., Kozák, P., *et al.* (2009) Prevalence of the crayfish plague pathogen *Aphanomyces astaci* in invasive American crayfishes in the Czech Republic. *Conservation Biology*, **23**, 1204–1213.

Kozubíková, E., Petrusek, A., Ďuriš, Z., *et al.* (2008) The old menace is back: Recent crayfish plague outbreaks in the Czech Republic. *Aquaculture*, **274**, 208–217.

Kozubíková, E., Viljamaa-Dirks, S., Heinikainen, S., *et al.* (2011) Spiny-cheek crayfish *Orconectes limosus* carry a novel genotype of the crayfish plague pathogen *Aphanomyces astaci*. *Journal of Invertebrate Pathology*, **108**, 214–216.

Kozubíková, E., Vrålstad, T., Filipová, L., *et al.* (2011) Re-examination of the prevalence of *Aphanomyces astaci* in North American crayfish populations in Central Europe by TaqMan MGB real-time PCR. *Diseases of Aquatic Organisms*, **97**, 113–125.

Kouba, A., Petrusek, A., Kozák, P.(2014) Continental-wide distribution of crayfish species in Europe: update and maps. *Knowledge and Management of Aquatic Ecosystems,* **413**,05.

Krueger, F. (2012) A wrapper tool around Cutadapt and FastQC to consistently apply quality and adapter trimming to FastQ files, with some extra functionality for MspI-digested RRBS-type (Reduced Representation Bisufite-Seq) libraries. *Available Online at: Http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/*.

Kumari, N., & Thakur, S. K. (2014) Randomly amplified polymorphic DNA-A brief review. *American Journal of Animal and Veterinary Sciences*, **9**, 6–13.

Kurtz, S., Phillippy, A., Delcher, A. L., *et al.* (2004) Versatile and open software for comparing large genomes. *Genome Biology*, **5**, R12.

Kušar, D., Vrezec, A., Ocepek, M., *et al.* (2013) *Aphanomyces astaci* in wild crayfish populations in Slovenia: first report of persistent infection in a stone crayfish *Austropotamobius torrentium* population. *Diseases of Aquatic Organinsms*, **103**, 157–169.

Lamour, K. H., Mudge, J., Gobena, D., *et al.* (2012) Genome sequencing and mapping reveal loss of heterozygosity as a mechanism for rapid adaptation in the vegetable pathogen *Phytophthora capsici*. *Molecular Plant-Microbe Interactions*, **25**, 1350–1360.

Leclerc, M. C., Guillot, J., & Deville, M. (2000) Taxonomic and phylogenetic analysis of Saprolegniaceae (Oomycetes) inferred from LSU rDNA and ITS sequence comparisons. *Antonie Van Leeuwenhoek*, **77**, 369–377.

Lee, S. B., & Taylor, J. W. (1992) Phylogeny of five fungus-like protoctistan *Phytophthora* species, inferred from the internal transcribed spacers of ribosomal DNA. *Molecular Biolology and Evolution*, **9**, 636–653.

Lévesque, C. A., Brouwer, H., Cano, L., *et al*. (2010) Genome sequence of the necrotrophic plant pathogen *Pythium ultimum* reveals original pathogenicity mechanisms and effector repertoire. Genome Biology, **11**, R73.

Lévesque, A. C., & De Cock, A. W. A. M. (2004) Molecular phylogeny and taxonomy of the genus *Pythium*. *Mycological Research*, **108**, 1363–1383.

Li, H. (2013) Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *Eprint arXiv:1303.3997v1*, **[q-GN]**.

Li, H., Handsaker, B., Wysoker, A., *et al.* (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, **25**, 2078–2079.

Li, X., Liu, Q.-H., & Huang, J. (2010) Effect of VP28 DNA vaccine on white spot syndrome virus in *Litopenaeus vannamei*. *Aquaculture International*, **18**, 1035–1044.

Lilley, J. H., Callinan, R. B., Chinabut, S., *et al.* (1998) *Epizootic Ulcerative Syndrome (EUS) Technical Handbook*. Bangkok: The Aquatic Animal

Health Research Institute.

Lilley, J. H., Cerenius, L., & Söderhäll, K. (1997a) RAPD evidence for the origin of crayfish plague outbreaks in Britain. *Aquaculture*, **157**, 181–185.

Lilley, J. H., Hart, D., Richards, R. H., *et al*. (1997b) Pan-Asian spread of single fungal clone results in larga scale fish kills. *Veterinary Record*, **140**, 653–654.

Liti, G., Carter, D. M., Moses, A. M., *et al.* (2009) Population genomics of domestic and wild yeasts. *Nature*, **458**, 337–341.

Little, T. J., O'Connor, B., Colegrave, N., *et al.* (2003) Maternal transfer of strain-specific immunity in an invertebrate. *Current Biology*, **13**, 489-492.

Luo, A., Zhang, A., Ho, S. Y. W., *et al.* (2011) Potential efficacy of mitochondrial genes for animal DNA barcoding: a case study using eutherian mammals. *BMC Genomics*, **12**, 1–13.

Luo, R., Liu, B., Xie, Y., *et al.* (2012) Software and supporting material for “SOAPdenovo2: an empirically improved memory-efficient short read de novo assembly." *GigaScience Database*.

Lymbery, A. J., Morine, M., Kanani, H. G., *et al.* (2014) Co-invaders: The effects of alien parasites on native hosts. *International Journal for Parasitology: Parasites and Wildlife*, **3**, 171–177.

Maguire, I., Jelić, M., Klobučar, G., *et al.* (2016) Prevalence of the pathogen *Aphanomyces astaci* in freshwater crayfish populations in Croatia. *Diseases of Aquatic Organinsms*, **118**, 45–53.

Makkonen, J., Jussila, J., Henttonen, P., *et al.* (2011) Genetic variation in the ribosomal internal transcribed spacers of *Aphanomyces astaci* Schikora from Finland. *Aquaculture*, **311**, 48–53.

Makkonen, J., Jussila, J., & Kokko, H. (2012) The diversity of the pathogenic Oomycete (*Aphanomyces astaci*) chitinase genes within the genotypes indicate adaptation to its hosts. *Fungal Genetics and Biology*, **49**, 635–642.

Makkonen, J., Jussila, J., Kortet, R., *et al.* (2012) Differing virulence of

Aphanomyces astaci isolates and elevated resistance of noble crayfish *Astacus astacus* against crayfish plague. *Diseases of Aquatic Organinsms*, **102**, 129–136.

Makkonen, J., Kokko, H., Vainikka, A., *et al.* (2014) Dose-dependent mortality of the noble crayfish (*Astacus astacus*) to different strains of the crayfish plague (*Aphanomyces astaci*). *Journal of Invertebrate Pathology*, **115**, 86–91.

Makkonen, J., Vesterbacka, A., Martin, F., *et al.* (2016) Mitochondrial genomes and comparative genomics of *Aphanomyces astaci* and *Aphanomyces invadans*. *Scientific Reports*, **6**, 36089.

Marino, F., Pretto, T., Tosi, F., *et al.* (2014) Mass Mortality of *Cherax quadricarinatus* (Von Martens, 1868) Reared in Sicily (Italy): Crayfish Plague Introduced in an Intensive Farming. *Freshwater Crayfish*, **20**, 93–96.

Martin, F. N., & Tooley, P. W. (2003) Phylogenetic relationships among *Phytophthora* species inferred from sequence analysis of mitochondrially encoded cytochrome oxidase I and II genes. *Mycologia*, **95**, 269–284.

Martin, M. (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal*, **17**, 10–12.

Martinati, P. (1862) Nota sulla malattia dei gamberi che ammorbò le acque del veronese nell'anno 1861. *Memorie Dell'accademia D'agricoltura Commercio Ed Arti Di Verona*, **XLI**, 215–226.

Martinez, D., Larrondo, L. F., Putnam, N., *et al.* (2004) Genome sequence of the lignocellulose degrading fungus *Phanerochaete chrysosporium* strain RP78. *Nature Biotechnology*, **22**, 695–700.

Mayr, E. (1942) Systematics and the origin of species. Columbia University Press, New York.

Mazzaglia, A., Studholme, D. J., Taratufolo, M. C., *et al.* (2012) *Pseudomonas syringae pv. actinidiae* (PSA) Isolates from Recent Bacterial Canker of Kiwifruit Outbreaks Belong to the Same Genetic Lineage. *PLoS One*, **7**,

e36518.

McDowell, J. M. (2011) Genomes of obligate plant pathogens reveal adaptations for obligate parasitism. *Proceedings of the National Academy of Sciences of the United States of America*, **108**, 8921–8922.

Metzker, M. L. (2005) Emerging technologies in DNA sequencing. *Genome Research*, **15**, 1767–1776.

Metzker, M. L. (2010) Sequencing technologies - the next generation. *Nature Reviews Genetics*, **11**, 31–46.

Michener, C. D. (1970) Diverse approaches to systematics. *Evolutionary Biology*, **4**, 1–38.

Misner, I., Blouin, N., Leonard, G., *et al.* (2015) The secreted proteins of *Achlya hypogyna* and *Thraustotheca clavata* identify the ancestral oomycete secretome and reveal gene acquisitions by horizontal gene transfer. *Genome Biology and Evolutio*n, **7**, 120–135.

Mrugała, A., Kawai, T., Kozubíková-Balcarová, E., *et al.* (2016) *Aphanomyces astaci* presence in Japan: a threat to the endemic and endangered crayfish species *Cambaroides japonicus*? *Aquatic Conservation: Marine and Freshwater Ecosystems*, **27**, 103–114.

Myers, E. W., Sutton, G. G., Delcher, A. L., *et al.* (2000) A whole-genome assembly of *Drosophila*. *Science*, **287**.

Nagarajan, N., & Pop, M. (2013) Sequence assembly demystified. *Nature Reviews Genetics*, **14**, 157–167.

Nielsen, R., Paul, J. S., Albrechtsen, A., *et al.* (2011) Genotype and SNP calling from next-generation sequencing data. *Nature Reviews Genetics,* **12**, 443–451.

Ninni, A. P. (1865) Sulla mortalita dei gamberi (*Astacus fluviatilis* L.) nel veneto e piu particolarmente nella provincia trevigiana. *Atti Instituto Veneto*, **10**, 1203–1209.

Nybelin, O. (1934) Nya undersökningar över kräftpestens orsak. *Ny Svensk*

Fiskeritidskrift, **1934**, 110–114.

Nyhlén, L., & Unestam, T. (1975) Ultrastructure of the penetration of the crayfish integument by the fungal parasite, *Aphanomyces astaci*, Oomycetes. *Journal of Invertebrate Pathology*, **26**, 353–366.

Oidtmann, B., Geiger, S., Steinbauer, P., *et al.* (2006) Detection of *Aphanomyces astaci* in North American crayfish by polymerase chain reaction. *Diseases of Aquatic Organinsms*, **72**, 53–64.

Oidtmann, B., Schaefers, N., Cerenius, L., *et al.* (2004) Detection of genomic DNA of the crayfish plague fungus *Aphanomyces astaci* (Oomycete) in clinical samples by PCR. *Veterinary Microbiology*, **100**, 269–282.

Oidtmann, B., Schmid, I., Rogers, D., *et al.* (1999) An improved isolation method for the cultivation of the crayfish plague fungus, *Aphanomyces astaci. Freshwater Crayfish*, **12**, 303–312.

OIE (2016a) Crayfish plague (*Aphanomyces astaci*). In: *World Organisation for Animal Health. Manual of Diagnostic Tests for Aquatic Animals.*, pp. 101–118.

OIE (2016b) Infection with *Aphanomyces invadans* (Epizootic Ulcerative Syndrome). In: *World Organisation for Animal Health. Manual of Diagnostic Tests for Aquatic Animals.*, pp. 1–14.

Olivera, I. E., Fins, K. C., Rodriguez, S. A., *et al.* (2016) Glycoside hydrolases family 20 (GH20) represent putative virulence factors that are shared by animal pathogenic oomycetes, but are absent in phytopathogens. *BMC Microbiology*, **16**, 232.

Onions, A. H. S. (1971) Preservation of fungi. In: *Methods in Microbiology* (ed Booth, C.), pp. 113–115. Academic Press, London and New York.

Paasonen, P., Edgerton, B., Tapiovaara, H., *et al.* (1999) *Freshwater crayfish virus research in Finland: state of the art*. Oslo.

Panabieres, F., Ponchet, M., Allasia, V., *et al*. (1997) Characterization of border species among Pythiaceae: several *Pythium* isolates produce elicitins, typical proteins from Phytophthora spp. *Mycological Research*, **101**,1459–

1468.

Parra, G., Bradnam, K., Ning, Z., *et al.* (2009) Assessing the gene space in draft genomes. *Nucleic Acids Research*, **37**, 289–297.

Pârvulescu, L., Schrimpf, A., Kozubíková, E., *et al.* (2012) Invasive crayfish and crayfish plague on the move: first detection of the plague agent *Aphanomyces astaci* in the Romanian Danube. *Diseases of Aquatic Organinsms*, **98**, 85–94.

Paszkiewicz, K., & Studholme, D. J. (2010) *De novo* assembly of short sequence reads. *Briefings in Bioinformatics*, **11**, 457–472.

Patoka, J., Kocánová, B., & Kalous, L. (2016) Crayfish in Czech cultural space: the longest documented relationship between humans and crayfish in Europe. *Knowledge and Management of Aquatic Ecosystems*, **417**, 5.

Pawlowska, T. E., & Taylor, J. W. (2004) Organization of genetic variation in individuals of arbuscular mycorrhizal fungi. *Nature*, **427**, 733–737.

Peeler, E. J., Oidtmann, B. C., Midtlyng, P. J., *et al.* (2011) Non-native aquatic animals introductions have driven disease emergence in Europe. *Biological Invasions*, **13**, 1291–1303.

Pendleton, S., Hanning, I., Biswas, D., *et al.* (2013) Evaluation of whole-genome sequencing as a genotyping tool for *Campylobacter jejuni* in comparison with pulsed-field gel electrophoresis and flaA typing. *Poultry Science*, **92**, 573–580.

Penner, G. A., Bush, A., Wise, R., *et al.* (1993) Reproducibility of random amplified polymorphic DNA (RAPD) analysis among laboratories. *PCR Methods and Applications*, **2**, 341–345.

Pérez-Jiménez, M., Besnard, G., Dorado, G., *et al.* (2013) Varietal tracing of virgin olive oils based on plastid DNA variation profiling. *PLoS One*, **8**, e70507.

Persson, M., & Söderhäll, K. (1983) *Pacifastacus leniusculus* (Dana) and its resistance to the parasitic fungus *Aphanomyces astaci* Schikora. *Freshwater Crayfish*, **5**, 292−298.

Petersen, A. B., & Rosendahl, S. (2000) Phylogeny of the *Peronosporomycetes* (Oomycota) based on partial sequences of the large ribosomal subunit (LSU rDNA). *Mycological Research*, **104**, 1295–1303.

Pham, L. N., Dionne, M. S., Shirasu-Hiza, M., *et al*. (2007) A specific primed immune response in *Drosophila* is dependent on phagocytes. *PLoS Pathogens*, 3, e26.

Phillippy, A. M., Schatz, M. C., & Pop, M. (2008) Genome assembly forensics: finding the elusive mis-assembly. *Genome Biology*, **9**, R55.

Phillips, A. J., Anderson, V. L., Robertson, E. J., *et al.* (2008) New insights into animal pathogenic oomycetes. *Trends in Microbiology*, **16**, 13–19.

Phumichai, C., Phumichai, T., & Wongkaew, A (2015). Novel chloroplast microsatellite (cpSSR) markers for genetic diversity assessment of cultivated and wild *Hevea* rubber. P*lant Molecular Biology Report*, **33**,1486–1498.

Polglase, J. L., & Alderman, D. J. (1984) Crayfish plague threatens UK stock. *Fish Farmer*, **7**, 16–17.

Porter, T. M., & Golding, G. B. (2012) Factors that affect large subunit ribosomal DNA amplicon sequencing studies of fungal communities: classification method, primer choice, and error. *PLoS One*, **7**, e35749.

Pretto, T., Tosi, F., Sandoval-Sierra, J. V, *et al.* (2014) Characterization of *Aphanomyces astaci* in white-clawed crayfish *Austropotamobius pallipes* from Northern Italy: considerations regarding a crayfish plague outbreak. In: *IAA & CSJ Joint International Conference on Crustacea*. Sapporo, Japan.

Püfer, K., Munch, K., Hellmann, I., *et al.* (2012) The bonobo genome compared with the chimpanzee and human genomes. *Nature*, **486**, 527–531.

Quinlan, A. R., & Hall, I. M. (2010) BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*, **26**, 841–842.

Quiroz Velasquez, P. F., Abiff, S. K., Fins, K. C., *et al.* (2014) Transcriptome analysis of the entomopathogenic oomycete *Lagenidium giganteum* reveals

putative virulence factors. *Applied Environmental Microbiology*, **80**, 6427–6436.

Raffaele, S., Win, J., Cano, L. M., *et al*. (2010) Analyses of genome architecture and gene expression reveal novel candidate virulence factors in the secretome of *Phytophthora infestans*. *BMC Genomics*, **11**, 637

Rasko, D. A., Webster, D. R., Sahl, J. W., *et al.* (2011) Origins of the *E. coli* Strain Causing an Outbreak of Hemolytic–Uremic Syndrome in Germany. *The New England Journal of Medicine*, **365**, 709–717.

Rennerfelt, E. (1936) *Untersuchungen über die Entwicklung und Biologie des Krebspestpilzes Aphanomyces astaci Schikora*. Drottningholm.

Reynolds, J., & Souty-Grosset, C. (2012) *Management of freshwater biodiversity: crayfish as bioindicators*. Cambridge University Press.

Reynolds, J., Souty-Grosset, C., & Richardson, A. (2013) Ecological roles of crayfish in freshwater and terrestrial habitats. *Freshwater Crayfish*, **19**, 197–218.

Rezinciuc, S., Galindo, J., Montserrat, J., *et al.* (2014) AFLP-PCR and RAPD-PCR evidences of the transmission of the pathogen *Aphanomyces astaci* (Oomycetes) to wild populations of European crayfish from the invasive crayfish species, *Procambarus clarkii*. *Fungal Biology*, **118**, 612–620.

Rezinciuc, S., Sandoval-Sierra, J. V., Oidtmann,B., *et al*., (2015) The Biology of Crayfish Plague Pathogen *Aphanomyces astaci*: Current Answers to Most Frequent Questions. In: *Freshwater Crayfish, Chapter: 9*, (eds Kawai, T., Faulkes, Z., Scholtz, G.), pp.182–204. CRC Press.

Ribeiro, F. J., Przybylski, D., Yin, S., *et al.* (2012) Finished bacterial genomes from shotgun sequence data. *Genome Research*, **22**, 2270–2277.

Riethmüller, A., Weiß, M., & Oberwinkler, F. (1999) Phylogenetic studies of Saprolegniomycetidae and related groups based on nuclear large subunit ribosomal DNA sequences. *Canadian Journal of Botany*, **77**, 1790–1800.

Rizzo, D. M., Garbelotto, M., Davidson, J. M., *et al.* (2002) *Phytophthora ramorum* as the cause of extensive mortality of *Quercus* spp. and

Lithocarpus densiflorus in California. *Plant Diseases*, **86**, 205– 214.

Robideau, G. P., De Cock, A. W., Coffey, M. D., *et al.* (2011) DNA barcoding of oomycetes with cytochrome c oxidase subunit I and internal transcribed spacer. *Molecular Ecology Resources*, **11**, 1002–1011.

Robinson, J. T., Thorvaldsdottir, H., Winckler, W., *et al.* (2011) Integrative genomics viewer. *Nature Biotechnology*, **29**, 24–26.

Roer, R., & Dillaman, R. (1984) The structure and calcification of the crustacean cuticle. *American Zoologist*, **24**, 893–909.

Roetzer, A., Diel, R., Kohl, T. A., *et al.* (2013) Whole Genome Sequencing versus Traditional Genotyping for Investigation of a *Mycobacterium tuberculosis* Outbreak: A Longitudinal Molecular Epidemiological Study. *PLoS Medicine*, **10**, e1001387.

Roth, O., Joop, G., Eggert, H., *et al*. (2010) Paternally derived priming for offspring in the red flour beetle, *Tribolium castaneum.* J*ournal of Animal Ecology*, **79**, 403-413.

Rowley, A. F., Pope, E. C. (2012) Vaccines and crustacean aquaculture—A mechanistic exploration. *Aquaculture*, **334-337**, 1–11.

Royo, F., Andersson, G., Bangyeekhun, E., *et al.* (2004) Physiological and genetic characterisation of some new *Aphanomyces* strains isolated from freshwater crayfish. *Veterinary Microbiology*, **104**, 103–112.

Rushton, S. P., Lurz, P. W. W., Gurnell, J., *et al.* (2006) Disease threats posed by alien species: the role of a poxvirus in the decline of the native red squirrel in Britain. *Epidemiology and Infection*, **134**, 521–533.

Sabat, A. J., Budimir, A., Nashev, D., *et al.* (2013) Overview of molecular typing methods for outbreak detection and epidemiological surveillance. *Euro Surveill*, **18**, 20380.

Sambrook, J., & Russell, D. W. (2012) Molecular cloning:a laboratory manual, 4th Edition. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

Sarathi, M., Simon, M. C., Ventatesan, C., *et al.* (2008) Oral administration of bacterially expressed VP28dsRNA to protect *Penaeus monodon* from white spot syndrome virus. *Marine Biotechnology*, **10**, 242–249.

Salzberg, S. L., & Yorke, J. A. (2005) Beware of mis-assembled genomes. *Bioinformatics*, **21**, 4320–4321.

Sanger, F., Donelson, J. E., Coulson, A. R., *et al.* (1973) Use of DNA Polymerase I Primed by a Synthetic Oligonucleotide to Determine a Nucleotide Sequence in Phage f1 DNA. *Proceedings of the National Academy of Sciences of the United States of America*, **70**, 1209–1213.

Schadt, E. E., Turner, S., & Kasarskis, A. (2010) A window into third-generation sequencing. *Human Molecular Genetics*, **19**, R227–R240.

Schindel, D. E., & Miller, S. E. (2005) DNA barcoding a useful tool for taxonomists. *Nature*, **435**, 17.

Schoch, C. L., Seifert, K. A., Huhndorf, S., *et al.* (2012) Nuclear ribosomal internal transcribed spacer (ITS) region as a universal DNA barcode marker for Fungi. *Proceedings of the Natural Academy of Science of the United States of America*, **109**, 6241–6246.

Schornack, S., van Damme, M., Bozkurt, T. O., *et al.* (2010) Ancient class of translocated oomycete effectors targets the host nucleus. *Proceedings of the National Academy of Sciences*, **107**, 17421–17426.

Shrestha, S., Hu, J., Fryxell, R. T., *et al.* (2014) SNP markers identify widely distributed clonal lineages of *Phytophthora colocasiae* in Vietnam, Hawaii and Hainan Island, China. *Mycologia*, **106**, 676–685.

Simão, F. A., Waterhouse, R. M., Ioannidis, P., *et al.* (2015) BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*, **31**, 3210–3212.

Simon, U. K., & Weiß, M. (2008) Intragenomic variation of fungal ribosomal genes is higher than previously thought. *Molecular Biology and Evolution*, **25**, 2251–2254.

Singh, I. S. B., Manjusha, M., Pai, S. S., *et al.* (2005). *Fenneropenaeus indicus*

is protected from white spot disease by oral administration of inactivated white spot syndrome virus. *Diseases of Aquatic Organisms*, **66**, 265–270.

Skroch, P., & Nienhuis, J. (1995) Impact of scoring error and reproducibility RAPD data on RAPD based estimates of genetic distance. *Theoretical and Applied Genetics*, **91**, 1086–1091.

Sneath, P. H. A., & Sokal, R. R. (1973) Numerical taxonomy: The principles and practice of numerical classification. W. H. Freeman, San Francisco.

Soanes, D. M., & Talbot, N. J. (2008) Moving targets: rapid evolution of oomycete effectors. *Trends in Microbiology*, **16**, 507–510.

Söderhäll, K., & Cerenius, L. (1999) The crayfish plague fungus: history and recent advances. *Freshwater Crayfish*, **12**, 11–35.

Söderhäll, K., Dick, M. W., Clark, G., *et al.* (1991) Isolation of *Saprolegnia parasitica* from the crayfish *Astacus leptodactylus*. *Aquaculture*, **92**, 121–125.

Söderhäll, K., Rantamäki, J., & Constantinescu, O. (1993) Isolation of *Trichosporon beigelii* from the freshwater crayfish *Astacus astacus*. *Aquaculture*, **116**, 25–31.

Souty-Grosset, C., Holdich, D. M., Noël, P. Y., *et al*. (eds.) (2006) A*tlas of Crayfish in Europe.* Muséum national d'Histoire naturelle, Paris, Patrimoines naturels, 64, 187 p.

Stajich, J. E., Wilke, S. K., Ahrén, D., *et al.* (2010) Insights into evolution of multicellular fungi from the assembled chromosomes of the mushroom *Coprinopsis cinerea* (*Coprinus cinereus*). *Proceedings of the National Academy of Sciences* , **107**, 11889–11894.

Stam, R., Jupe, J., Howden, A. J. M, *et al.* (2013) Identification and characterisation CRN effectors in *Phytophthora capsici* shows modularity and functional diversity. *PLOS One,* **8**,

Studholme, D. J. (2016) Genome Update. Let the consumer beware: Streptomyces genome sequence quality. *Microbial Biotechnology*, **9**, 3–7.

Studholme, D. J., Glover, R. H., & Boonham, N. (2011) Application of high-throughput DNA sequencing in phytopathology. *Annual Review of Phytopathology*, **49**, 87–105.

Svoboda, J., Mrugała, A., Kozubíková-Balcarová, E., *et al.* (2017) Hosts and transmission of the crayfish plague pathogen *Aphanomyces astaci*: a review. *Journal of Fish Diseases*, **40**, 127–140.

Takuma, D., Sano, A., & Hatai, K. (2013) Two new species, *Aphanomyces izumoensis* sp. nov. and *Aphanomyces shimanensis* sp. nov. isolated from Ice Fish *Salangichthys microdon*. *International Journal of Research in Pure and Applied Microbiology*, **3**, 67–76.

Tamura, K., & Nei, M. (1993) Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Molecular Biology and Evolution*, **10**, 512–526.

Tamura, K., Stecher, G., Peterson, D., *et al.* (2013) MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Molecular Biology and Evolution*, **30**, 2725–2729.

Taugbøl, T., Skurdal, J., & Håstein, T. (1993) Crayfish plague and management strategies in Norway. *Biological Conservation*, **63**, 75–82.

Taylor, P., & Lewontin, R. (2017) The Genotype/Phenotype Distinction. T*he Stanford Encyclopedia of Philosophy* (ed Zalta, N. Z.). https://plato.stanford.edu/entries/genotype-phenotype/

The Chimpanzee Sequencing and Analysis Consortium (2005) Initial sequence of the chimpanzee genome and comparison with the human genome. *Nature*, **437**, 69-87.

Thines, M., & Kamoun, S. (2010) Oomycete–plant coevolution: recent advances and future prospects. *Current Opinion in Plant Biology*, **13**, 427–433.

Thörnqvist, P.-O., & Söderhäll, K. (1993) *Psorospermium haeckeli* and its interaction with the crayfish defence system. *Aquaculture*, **117**, 205–213.

Thorvaldsdóttir, H., Robinson, J. T., & Mesirov, J. P. (2013) Integrative Genomics Viewer (IGV): high-performance genomics data visualization and

exploration. *Briefings in Bioinformatics*, **14**, 178–192.

Tiffney, W. N. (1939) The host range of *Saprolegnia parasitica*. *Mycologia*, **31**, 310–321.

Tingey, S. V, & del Tufo, J. P. (1993) Genetic analysis with random amplified polymorphic DNA markers. *Plant Physiology*, **101**, 349–352.

Tuffs, S., & Oidtmann, B. (2011) A comparative study of molecular diagnostic methods designed to detect the crayfish plague pathogen, *Aphanomyces astaci*. *Veterinary Microbiology*, **153**, 343–353.

Turner, T. L, Hahn, M. W., & Nuzhdin, S. V. (2005). Genomic islands of speciation in *Anopheles gambiae. PLoS Biology,* **3**, e285.

Unestam, T. (1972) On the host range and origin of the crayfish plague fungus. *Reports from the Institute of Freshwater Research, Drottningholm*, **52**, 192–198.

Unestam, T., & Weiss, D. W. (1970) The Host-Parasite Relationship between Freshwater Crayfish and the Crayfish Disease Fungus *Aphanomyces astaci*: Responses to Infection by a Susceptible and a Resistant Species. *Journal of General Microbiology*, **60**, 77–90.

van den Berg, A.H., McLaggan, D., Diéguez-Uribeondo, J., *et al*. (2013). The impact of the water moulds *Saprolegnia diclina* and *Saprolegnia parasitica* on natural ecosystems and the aquaculture industry. *Fungal Biology Reviews*, **27**, 33–42.

Van West, P., de Bruijn, I., Minor, K. L., *et al.* (2010) The putative RxLR effector protein SpHtp1 from the fish pathogenic oomycete *Saprolegnia parasitica* is translocated into fish cells. *FEMS Microbiology,* **310**, 127–137.

Varki, A., & Altheide, T. K. (2005) Comparing the human and chimpanzee genomes: searching for needles in a haystack. *Genome Research*, **15**, 1746-1758.

Viljamaa-Dirks, S., Heinikainen, S., Nieminen, M., *et al.* (2011) Persistent infection by crayfish plague *Aphanomyces astaci* in a noble crayfish population – a case report. *Bulletin of the European Association of Fish*

Pathologists, **31**, 182–188.

Viljamaa-Dirks, S., Heinikainen, S., Torssonen, H., *et al.* (2013) Distribution and epidemiology of genotypes of the crayfish plague agent *Aphanomyces astaci* from noble crayfish *Astacus astacus* in Finland. *Diseases of Aquatic Organinsms*, **103**, 199–208.

Viljamaa-Dirks, S., Heinikainen, S., Virtala, A. M. K., *et al.* (2016) Variation in the hyphal growth rate and the virulence of two genotypes of the crayfish plague organism *Aphanomyces astaci*. *Journal of Fish Diseases*, **39**, 753–764.

Viljamaa-Dirks, S., Mrugała, A., Kozubíková-Balcarová, E., *et al.* (2014) Crayfish plague isolates from ornamental crayfish trade show genotypic variation. In: *IAA & CSJ Joint International Conference on Crustacea*. Sapporo, Japan.

Vos, P., Hogers, R., Bleeker, M., *et al.* (1995) AFLP: a new technique for DNA fingerprinting. *Nucleic Acids Research*, **23**, 4407–4414.

Vrålstad, T., Johnsen, S. I., Fristad, R. F., *et al.* (2011) Potent infection reservoir of crayfish plague now permanently established in Norway. *Diseases of Aquatic Organisms*, **97**, 75–83.

Vrålstad, T., Johnsen, S. I., & Taugbøl, T. (2011) NOBANIS – Invasive Alien Species Fact Sheet – *Aphanomyces astaci*. – From: Online Database of the European Network on Invasive Alien Species – NOBANIS www.nobanis.org, Date of access 29/06/2017.

Vrålstad, T., Knutsen, A. K., Tengs, T., *et al.* (2009) A quantitative TaqMan{®} MGB real-time polymerase chain reaction based assay for detection of the causative agent of crayfish plague *Aphanomyces astaci. Veterinary Microbiology*, **137**, 146–155.

Vrålstad, T., Strand, D. A., Grandjean, F., *et al.* (2014) Molecular detection and genotyping of *Aphanomyces astaci* directly from preserved crayfish samples uncovers the Norwegian crayfish plague disease history. *Veterinary Microbiology*, **173**, 66–75.

Wagner, A., Blackstone, N., Cartwright, P., *et al.* (1994) Surveys of gene families using polymerase chain reaction: PCR selection and PCR drift. *Systematic Biology*, **43**, 250–261.

Wang, X., Wang, H., Wang, J., *et al.* (2011) The genome of the mesopolyploid crop species *Brassica rapa*. *Nature Genetics*, **43**, 1035–1039.

Wang, Z., Chen, Y., & Li, Y. (2004) A Brief Review of Computational Gene Prediction Methods. *Genomics, Proteomics and Bioinformatics*, **2**, 216–221.

Waugh, J. (2007) DNA barcoding in animal species: progress, potential and pitfalls. *Bioessays*, **29**, 188–197.

Wawra, S., Bain, J., Durward, E., *et al.* (2012) Host-targeting protein 1 (SpHtp1) from the oomycete *Saprolegnia parasitica* translocates specifically into fish cells in a tyrosine-O-sulphate-dependent manner. *PNAS*, **109**, 2096–2101.

Wei, K.-Q., & Xu, Z.-R. (2009) Effects of oral recombinant VP28 expressed in silkworm (*Bombyx mori*) pupa on immune response and disease resistance of *Procambarus clarkii. World Journal of Microbiology and Biotechnology,* **25**, 1321–1328.

Whisson, S. C., Boevink, P. C., Moleleki, L. *et al.* (2007) A translocation signal for delivery of oomycete effector proteins into host plant cells. *Nature*, **450**, 115–118.

White, T. J., Bruns, T., Lee, S., *et al.* (1990) Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. In: *PRC Protocols: A Guide to Methods and Applications* (eds Innis, M. A., Gelfand, D. H., Sninsky, J. J., *et al*.), p. 482. New York: Academic Pres, Inc.

Willoughby, L. G., & Roberts, R. J. (1994) Improved methodology for isolation of the *Aphanomyces* fungal pathogen of epizootic ulcerative syndrome (EUS) in Asian fish. *Journal of Fish Diseases,* **17**, 541–543.

Wilkerson, R. C., Parsons, T. J., Albright, D. G., *et al.* (1993) Random amplified polymorphic DNA (RAPD) markers readily distinguish cryptic mosquito species (Diptera: Culicidae: *Anopheles*). *Insect Molecular Biology*, **1**, 205–

211.

Williams, J. G., Kubelik, A. R., Livak, K. J., *et al.* (1990) DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. *Nucleic Acids Research*, **18**, 6531–6535.

Wilson, I. G. (1997) Inhibition and facilitation of nucleic acid amplification. *Applied and Environmental Microbiology*, **63**, 3741–3751.

Win, J., & Kamoun, S. (2008) Adaptive evolution has targeted the C-terminal domain of the RXLR effectors of plant pathogenic oomycetes. *Plant Signaling & Behavior*, **3**, 251–253.

Wright, S. (1940) The statistical consequences of Mendelian heredity in relation to speciation. In: *The new systematics* (ed Huxley, J.), pp 161–183. Oxford University Press, London.

Zerbino, D. R., & Birney, E. (2008) Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Research*, **18**, 821–829.